Journal of Mathematical Biology (2023) 87:87
https://doi.org/10.1007/s00285-023-01995-5

Mathematical Biology

A modeling framework for adaptive collective defense: crisis response in social-insect colonies

M. Gabriela Navas-Zuloaga[1] · Kaitlin M. Baudier[2] · Jennifer H. Fewell[3] · Noam Ben-Asher[4] · Theodore P. Pavlic[3,5,6,7] · Yun Kang[8]

Received: 25 October 2022 / Revised: 26 August 2023 / Accepted: 7 September 2023 /
Published online: 15 November 2023

Abstract

Living systems, from cells to superorganismic insect colonies, have an organizational boundary between inside and outside and allocate resources to defend it. Whereas the micro-scale dynamics of cell walls can be difficult to study, the adaptive allocation of workers to defense in social-insect colonies is more conspicuous. This is particularly the case for *Tetragonisca angustula* stingless bees, which combine different defensive mechanisms found across other colonial animals: (1) morphological specialization (distinct soldiers (majors) are produced over weeks); (2) age-based polyethism (young majors transition to guarding tasks over days); and (3) task switching (small workers (minors) replace soldiers within minutes under crisis). To better understand how these timescales of reproduction, development, and behavior integrate to balance

✉ M. Gabriela Navas-Zuloaga
mnavaszu@asu.edu

✉ Yun Kang
ykang3@asu.edu

1 School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA

2 School of Biological, Environmental, and Earth Sciences, The University of Southern Mississippi, Hattiesburg, MS 39406, USA

3 School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA

4 Data Science Directorate, SimSpace Cooperation, Boston, MA, USA

5 School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA

6 School of Sustainability, Arizona State University, Tempe, AZ 85281, USA

7 School of Complex Adaptive Systems, Arizona State University, Tempe, AZ 85281, USA

8 Sciences and Mathematics Faculty, College of Integrative Sciences and Arts, Arizona State University, Tempe, AZ 85281, USA

defensive demands with other colony needs, we developed a demographic Filippov ODE system to study the effect of these processes on task allocation and colony size. Our results show that colony size peaks at low proportions of majors, but colonies die if minors are too plastic or defensive demands are too high or if there is a high proportion of quickly developing majors. For fast maturation, increasing major production may decrease defenses. This model elucidates the demographic factors constraining collective defense regulation in social insects while also suggesting new explanations for variation in defensive allocation at smaller scales where the mechanisms underlying defensive processes are not easily observable. Moreover, our work helps to establish social insects as model organisms for understanding other systems where the transaction costs for component turnover are nontrivial, as in manufacturing systems and just-in-time supply chains.

Keywords Mathematical biology · Social insects · Task allocation · Collective behavior · Collective defense

Mathematics Subject Classification 92-10

1 Introduction

Efficiently allocating resources to tasks is an important function for any biological organism. At a fundamental level, resources necessary for biological function are both limited and have a bounded replenishment rate (Caetano-Anollés et al. 2021). Consequently, there are significant opportunity costs concomitant with any over-supply of resources to a task. Furthermore, resource deficits can spontaneously occur with previously balanced allocations if surrounding conditions suddenly change. These potential problems are amplified when allocated resources are inflexible or unable to be quickly reallocated from one task to another. Eusocial-insect colonies are valuable models to study these fundamental problems as their size (relative to other biological collectives) allows individuals to be directly observed and they can exhibit high degrees of diversity in worker specialization and must shape worker demographics in a completely decentralized manner (Camazine et al. 2003; Gordon 2002; Beshers and Fewell 2001). Intricate adaptive colony dynamics and sophisticated division of labor emerge exclusively from local interactions between individuals without global information and result in a highly complex, distributed organization that is in certain ways superior to hierarchical organizations with central control (Fewell 2003; Holbrook et al. 2009).

The efficient allocation of finite resources to the task of colony defense is an important problem to eusocial insects, which accumulate resources internally that can become targets for kleptoparasitic robbing and raiding behavior by other insects (Baudier et al. 2019). Insect societies in general often exhibit some degree of specialization; for example, tasks are commonly determined by age, which is known as age-based polyethism and corresponds to specialization inside an age group (Baudier et al. 2019; Seeley 1982). As with any other task, increased use of specialized workers for defense is subject to a flexibility–specialization trade-off. On the one hand, specialization can increase group-level energetic efficiency in the long term (Jeanne

1986), but it also slows reactions to sudden changes (Dornhaus 2008), like unexpected threats (e.g., large raids from other colonies). For the case of social insects, extreme specialization takes the form of morphologically distinct individuals that perform a narrow range of tasks very efficiently but can be costly to produce. On the other hand, individuals that can flexibly switch between tasks according to demand allow a colony to quickly adapt to emergency needs, manifested in acute changes in task demand (Dornhaus 2008; Jongepier and Foitzik 2016), but these flexible generalists are individually less capable for some or all of the tasks in their repertoire. In this work, we examine how a hierarchy of mechanisms in eusocial-insect colonies can regulate task allocation to maintain colony growth and reproduction while dedicating sufficient workforce to group defense in response to emergent threats posed by their dynamic environment.

The stingless bee *Tetragonisca angustula* is an ideal model organism to study the flexibility–specialization trade-off in collective defense as colonies employ multiple task-allocation mechanisms ranging over different timescales and degrees of flexibility (Baudier et al. 2019). Although worker morphological specialization is very rare among bees, *T. angustula* colonies produce a minority of large-bodied workers (*majors*) that are more efficient at nest defense than their smaller nestmates (*minors*) but also require more resources to be produced (Grüter et al. 2012; Jones et al. 2012; van Zweden et al. 2011). The size and future developmental trajectory of adult bees at eclosion (i.e., emerging in adult form from the pupal case) depends on their earlier feeding schedule while larval brood, which in turn is determined by their rearing location and the feeding behavior of nurse bees (Segers et al. 2015). Moreover, colonies exposed to a higher frequency of threats produce a larger proportion of majors (Segers et al. 2016), likely due to adaptive changes in how those brood are reared in response to the increased threat. In addition to morphological specialization, age-based polyethism is also present in the form of age-dependent task allocation. Young workers of all sizes perform mostly brood care and nest maintenance whereas older bees (about two weeks old) tend to work outside the nest, either foraging in the case of minors or guarding in the case of majors (Hammel et al. 2016). While age-based polyethism provides the colony some degree of flexibility, crisis situations that require a fast response are addressed through a third mechanism involving behavioral plasticity of minors. In response to guard loss, minor workers replace guards within minutes (Baudier et al. 2019), which provides a temporary defensive reinforcement to the colony during the relatively longer developmental period while new major guards are produced.

Several quantitative models of task allocation have been studied for eusocial insects (Beshers and Fewell 2001), but very little work has focused on the balance between colony growth and defense mechanisms. Kang and Theraulaz (2016) developed a multi-compartment differential-equation model for division of labor in social insect colonies that, although not focusing on defense, included age-based polyethism and task switching as mechanisms regulating worker allocation. Their model does not account for morphological specialization and assumes that all tasks contribute equally to the eclosion of new adults. In a different study, Aoki and Kurosu (2003) explicitly modeled the production of morphologically distinct, non-reproductive soldiers based on their productivity for the colony. That model does focus on the balance between defense and reproduction, but the chosen organism has no allocation flexibility in terms of age-based polyethism or task switching. Recent work by Strickland et al. (2019) uses stingless bees as inspiration to develop an algorithm for allocation of guarding tasks in robots. Their study explores the value of a heterogeneous guarding force with different defense tasks for both robotic swarms and stingless bees in terms of performance, but it is not concerned with group growth or reproduction.

In order to better understand how social insects regulate colony growth and defense efficiently, we have designed a demographic model of task allocation in *T. angustula* colonies. Our proposed model uses a system of differential equations that explicitly reflects processes occurring at three distinct timescales (i.e., morphological specialization, age-based polyethism, and behavioral plasticity) in relation to colony defense and growth. We use the model to study plausible conditions under which the studied defense mechanisms work in tandem to improve collective defense. Specifically, we address the following questions:

1. How do the parameters regulating morphological specialization, age-based polyethism, and behavioral plasticity impact the colony size and task allocation within the colony?
2. What degree of behavioral plasticity allows the colony to transiently compensate for loss of defenses without maintaining an unnecessary population of inefficient minor guards?

The remainder of the paper is organized as follows. In Sect. 2, we derive an ordinary-differential-equation model to describe the population dynamics and task allocation within *T. angustula* stingless bee colonies. In Sect. 3, we provide a theoretical analysis of the model's dynamical properties. In Sect. 4, we study the interaction between morphological specialization, age-based polyethism, and behavioral plasticity for colony survival, growth and task allocation through bifurcation analysis and simulations in biologically realistic scenarios. Lastly, we include concluding remarks and future work in Sect. 5.

2 Model derivation

In this section, we derive a dynamical model of task allocation in stingless bee colonies that includes: (1) morphological specialization, (2) age-based polyethism, and (3) task switching.

2.1 Worker types and task types

We assume that all adult bees in a colony are either large-bodied majors (i.e., specialized for defense tasks) or smaller minors according to their body size at eclosion. This accounts for the morphological specialization observed by Grüter et al. (2012) and Segers et al. (2015). Because our focus is defense regulation, we consider only two types of tasks that majors and minors may perform: guarding or non-guarding. Both majors and minors begin their lives doing non-guarding tasks, but majors mature twice as fast as minors and spend the last half of their lives as defensive "soldiers" in guarding tasks (*guards* herein). However, minors only transition to guarding when there are

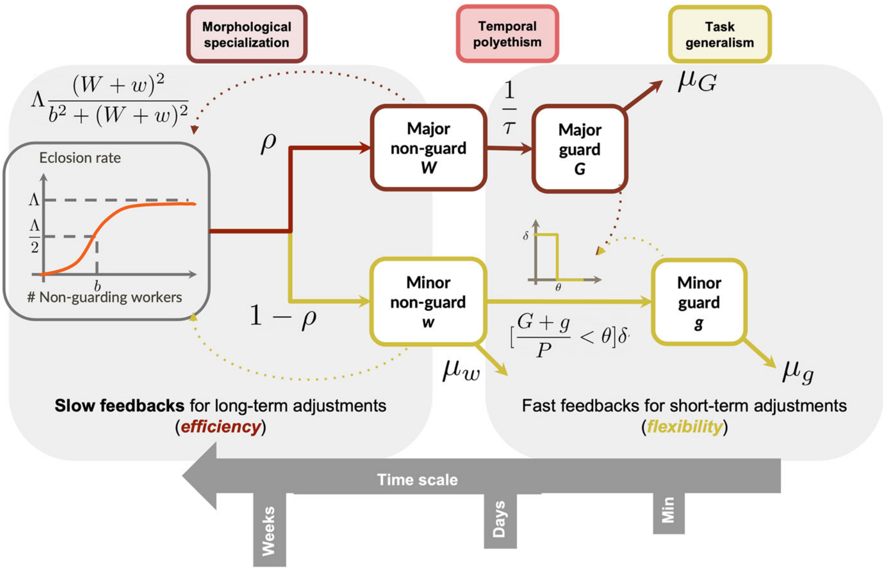

Fig. 1 Model diagram. Eggs are laid at a constant rate Λ by the queen and a fraction of those, dependent on the non-guarding population, successfully eclose as adults (see text for more details). A percent ρ of the newborns are majors, the rest are minors. Young majors are workers inside the nest and transition to guarding tasks after a maturation time τ. Then they die at a rate μ_G. Minors usually perform non-guarding tasks and die at a rate μ_w, but if the proportion of guards in the colony drops below a threshold θ, they replace them as guards at a rate δ. Minor guards die at a rate μ_g. The Iverson (1962) bracket $[\cdot]$ is 1 when its predicate argument is true and 0 otherwise

sudden deficits in guarding and otherwise perform only non-guarding tasks. Thus, at any time, all adult individuals can be divided among four possible compartments:

- Major Non-guard (W): large-bodied bee performing non-guarding tasks
- Major Guard (G): large-bodied bee performing guarding tasks
- Minor Non-guard (w): smaller-bodied bee performing non-guarding tasks
- Minor Guard (g) smaller-bodied bee performing guarding tasks

Based on context, the variables (W, G, w, g) will refer to either the number of adult bees in each compartment or the name of the compartment itself. With this nomenclature, the population size P, subject to demographic changes, can be written as

$$P := W + G + w + g.$$

Next, we describe the dynamic fluxes into and out of these four compartments, which are summarized in Fig. 1.

2.2 Production of new major and minor workers

The queen lays eggs at an average rate Λ; however, we assume that the $G + g$ adult bees performing guarding tasks do not contribute to brood care. This assumption is based on the experimental observation that guards (G and g) mainly patrol, hover, or stand near the nest entrance (Hammel et al. 2016). So the fraction of eggs that successfully eclose to adults depends on the size of the working non-guard population, $W + w$. Consequently, a colony consisting entirely of guards (i.e., $P = G + g$) and no workers (i.e., $W = w = 0$) will not have a positive growth rate in this model, and so there is a trade-off between growth and defense that manifests through devoting resources to guards or workers. To describe the production of new adults, we choose the commonly used Hill form (Goutelle et al. 2008):

$$\Lambda \frac{(W+w)^2}{(W+w)^2 + b^2} = \Lambda \frac{1}{1 + \left(\dfrac{b}{W+w}\right)^2} \tag{1}$$

where the half-saturation constant b represents the number of non-guarding workers ($W + w$) required for the eclosion rate to be half of the egg-laying rate Λ (Kang and Theraulaz 2016; Kang et al. 2011, 2016; Eberl et al. 2010; Ratti et al. 2012; Britton and White 2021). The Hill exponent of 2 provides a type-III functional response where eclosion rate accelerates at low numbers of non-guard workers (i.e., when there is a high demand for nurses and foragers) but then has diminishing marginal returns at high numbers of workers (i.e., as brood-rearing demand is met). The value of the exponent is positively correlated with the number of tasks or sets of tasks required to maintain the brood; in this case, it represents the interaction between the two broad sets of in-nest (e.g. brood care) and outside (e.g. foraging) tasks (Kang and Theraulaz 2016; Kang et al. 2011).

Experimental studies have shown that the size of the adult bees is determined before eclosion by the feeding schedule that they experience as larvae (Segers et al. 2015) and that colonies exhibit two distinct body sizes corresponding to major and minor bees (Grüter et al. 2012). Guided by these facts, we assign a proportion $\rho \in (0, 1)$ of the newly emerged adults to be majors (large-bodied bees, specialized for defense) and the remaining $1 - \rho$ to be minors (smaller-bodied bees, primarily taking part in non-guarding tasks).

Evidence suggests that colonies can regulate the proportion of majors produced in the long term according to environmental threats (Segers et al. 2016). However, bees need approximately 5–6 weeks from egg to eclosion plus an additional 2 weeks of maturation to become guards, which means that rearing a new generation of guards requires 7–8 weeks (Segers et al. 2016). Thus, changes in the proportion of majors take approximately 2 bee lifespans to have an effect. In this model, we assume a constant proportion ρ and manipulate this parameter to observe its effect on system trajectories.

2.3 Developmental dynamics of majors

Majors perform in-nest tasks or foraging when they are young and transition to guarding when they reach middle age (Hammel et al. 2016). Once they start guarding, going back to non-guarding tasks is rare (Baudier et al. 2019). Recall that W and G represent the number of majors performing non-guarding and guarding tasks, respectively; we let τ be the average W-to-G maturation time for majors to start guarding after eclosion. Because majors transition into guarding at a relatively young age, we have omitted a mortality rate for major non-guards for simplicity. Mature major guards (G) die at a rate μ_G (see Fig. 1). Thus, the life of majors is modeled by the system:

$$\begin{cases} \dfrac{\mathrm{d}W}{\mathrm{d}t} = \underbrace{\Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}\rho}_{\text{eclosion}} \underbrace{-\dfrac{1}{\tau}W}_{\text{maturation}} \\ \dfrac{\mathrm{d}G}{\mathrm{d}t} = \underbrace{\dfrac{1}{\tau}W}_{\text{maturation}} \underbrace{-\mu_G G}_{\text{death}} \end{cases} \tag{2}$$

The time-constant parameter τ implements age-based polyethism in the model, which is the correlation of age and task within insect colonies. It represents the average time needed for a worker to develop into a guard, and thus it delays the effect of changing the proportion ρ of majors produced on colony defense. We will study how τ interacts with ρ in determining the availability of guards for the colony and how these two parameters relate to the process of guard replacement by minors in case of emergency.

2.4 Developmental dynamics of minors with behavioral switching

Under normal conditions, minor workers perform in-nest tasks during young and middle age and switch to foraging when they reach old age (Hammel et al. 2016). We denote by w the number of working minors in all of those tasks and μ_w their average mortality rate. The nominal life of a minor worker is therefore represented by:

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \underbrace{\Lambda \frac{(W+w)^2}{b^2+(W+w)^2}(1-\rho)}_{\text{eclosion}} \underbrace{-\mu_w w}_{\text{death}}. \tag{3}$$

However, in the absence of guards G (e.g., due to a crisis where guards are lost after a raid by an invading colony), minors adopt guarding behavior. When guards were removed from a natural colony by Baudier et al. (2019), tracked minors were observed replacing guards just a few hours later. Thus, non-guarding minors w can transition to guarding minors g under the right conditions. To model this behavioral switch in minors, we assume that replacement occurs only when the total number of guards $G+g$ in the colony drops below a fraction θ of the total population P (i.e., when $(G+g)/P < \theta$). This characteristic models the scaling of guard demand with colony size. Furthermore, eusocial insects are known to exhibit density-dependent behaviors that are responsive to changes in local encounter rate (Gordon et al. 1993; Gordon and Mehdiabadi 1999; Pratt 2005; Gordon et al. 2008; Farji-Brener et al. 2010; Baudier and Pavlic 2020). Consequently, the guard density $(G+g)/P$ may be able to be inferred by individual workers using only local encounter information. Thus, we assume that the rate of replacement is a function of the per-capita probability of finding a guard among all nestmates ($(G+g)/P$). Based on the typical fraction of guards in wild colonies (Grüter et al. 2012; Hammel et al. 2016), the nominal value of this ratio is between 1 and 6%; we assume that workers have the ability to detect drops in guard ratio to below this value.

For simplicity, and to emphasize the difference of timescales between the behavioral replacement and developmental maturation processes, we also assume that replacement is binary and instantaneous: it is non-existent while $\dfrac{G+g}{P} > \theta$ and happens at a rate δ when $\dfrac{G+g}{P} \leq \theta$. So the behavioral w-to-g replacement rate is:

$$\delta\left[\frac{G+g}{P} \leq \theta\right] = \begin{cases} \delta, & \dfrac{G+g}{P} \leq \theta \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

where $[\cdot] : \{T, F\} \mapsto \{0, 1\}$ denotes the Iverson bracket (Iverson 1962). Note that replacement itself is not instantaneous for all minors, but rather occurs at a continuous rate δ. The assumption here is that minors can instantly sense a loss of guards below a critical threshold and become available for replacement only during the crisis. Thus, we model replacement not only as a "faster" process than maturation in the sense that $\delta > 1/\tau$, but also as a faster response to crisis in the sense that sudden guard loss triggers an immediate change in behavior (modeled by the Heaviside function), but not in development.

Combining (3) and (4), the general population dynamics of minor workers (guarding and non-guarding) facing potential crisis situations is the Variable Structure System (VSS) (Young and Özgüner 1999):

$$\begin{cases} \dfrac{\mathrm{d}w}{\mathrm{d}t} = \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}(1-\rho) - \mu_w w \underbrace{-\left[\dfrac{G+g}{P} \leq \theta\right]\delta w}_{\text{replacement}} \\ \\ \dfrac{\mathrm{d}g}{\mathrm{d}t} = \underbrace{\left[\dfrac{G+g}{P} \leq \theta\right]\delta w}_{\text{replacement}} \underbrace{-\mu_g g}_{\text{death}} \end{cases}. \tag{5}$$

In this system, the switching behavior acts as a bang–bang control (Bellman et al. 1956) on w-to-g replacement so as to regulate the guard ratio $(G+g)/P$ toward threshold parameter $\theta \in (0, 1)$. The magnitude of this threshold parameter represents how tolerant minors are to guard loss. In the ecologically relevant case where minor workers are responsible for most in-nest non-guarding work, the value of θ corresponds to an inherent trade-off in task allocation. If minors are too tolerant to guard loss (low θ), their failure to replace lost guards may leave the colony vulnerable for a long time before the next generation of guards is mature. However, for high values of θ, minors switch to guarding tasks even when a relatively high proportion of the colony is already guarding, and so they will not contribute much to defense and will instead fail to fulfill other tasks needed by the colony. In the coming sections, we will study how the replacement threshold (θ) interacts with the slower processes of maturation and production of majors to keep colony defense and growth balanced.

2.5 Full model

Combining (2) and (5) results in an autonomous ODE VSS that governs the general population dynamics of the colony:

$$\begin{cases} \dfrac{\mathrm{d}W}{\mathrm{d}t} = \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}\rho - \dfrac{1}{\tau}W \\ \dfrac{\mathrm{d}G}{\mathrm{d}t} = \dfrac{1}{\tau}W - \mu_G G \\ \dfrac{\mathrm{d}w}{\mathrm{d}t} = \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}(1-\rho) - \mu_w w - \left[\dfrac{G+g}{P} \leq \theta\right]\delta w \\ \dfrac{\mathrm{d}g}{\mathrm{d}t} = \left[\dfrac{G+g}{P} \leq \theta\right]\delta w - \mu_g g, \end{cases} \tag{6}$$

where a complete list of variables and parameters is summarized in Table 1. The reference parameter values that will be used throughout the analysis are in Table 2. Most values are taken from the literature from *T. angustula* stingless be colonies, but some are not found directly and must therefore be assumed. More details can be found in Appendix C. The proposed model incorporates all three mechanisms involved in colony defense by *T. angustula* stingless bees: production of major guards, adjustable over generations and mediated by the fraction ρ; maturation of such majors from non-guarding tasks to defensive guarding tasks (age-based polyethism), reflected in the maturation time τ; and quick replacement of guards by minor bees in case of

Table 1 Variable and parameter definition

Name	Definition	Units
State variables		
W	Number of non-guarding majors	individuals
G	Number of guard majors	individuals
w	Number of non-guarding minors	individuals
g	Number of guard minors	individuals
P	Total population size	individuals
Parameters		
Λ	Egg-laying rate of queen	individuals/day
b	Number of non-guarding workers for $\Lambda/2$ eclosion rate	individuals
ρ	Fraction of newborn majors	–
τ	Maturation time of bees	days
μ_G	Death rate of guarding majors	1/days
μ_g	Death rate of guarding minors	1/days
μ_w	Death rate of non-guarding minors	1/days
δ	Guard replacement rate for minors	1/days
θ	Replacement threshold for fraction of guards	-

Table 2 Parameter values for *T. angustula* stingless bees

Name	Value/range	Source
Parameters		
Λ	154 individuals/day	Koedam et al. (1997)
b	700–1300	Assumed
ρ	1–6%	Hammel et al. (2016)
τ	20 days	Hammel et al. (2016)
μ_G	$\left[\frac{1}{6.9}, \frac{1}{2.9}\right]$ days^{-1}	Grüter et al. (2011)
μ_g	$\left[\frac{1}{6.9}, \frac{1}{2.9}\right]$ days^{-1}	Grüter et al. (2011)
μ_w	$\left[\frac{1}{35}, \frac{1}{20}\right]$ days^{-1}	Grüter et al. (2011); Hammel et al. (2016)
δ	≥ 4 days^{-1}	Baudier et al. (2019)
θ	0.2–1.2%	Assumed

For Λ, we use the average value of 6.41 eggs laid per hour reported by Koedam et al. (1997) and multiply by 24 h to get an approximate daily rate. For b, we know that newly founded colonies have about 500–1000 workers (Van Veen and Sommeijer 2000) and we assume the Allee threshold E_{-}^{b} must be well below this range. We take an interval around b=1000, which, for $\Lambda = 154$, $\tau = 20$, $\mu_G = 1/5.4$, $\mu_g = 1/3$, $\mu_w = 1/28$, $\delta = 4$, $\rho = 0.06$, yields an Allee threshold of 250 individuals. For θ, we assume that it must be below γ_b for natural colonies because they have no minor guards. With the same parameter values, we calculated the range for θ corresponding to ρ between 0.01 and 0.06

emergency, where replacement rate δ is triggered when the guard ratio falls below threshold θ. To our knowledge, this is the first mathematical model to include all three mechanisms in relation to colony defense in polymorphic eusocial insects, and specifically in bees, where morphological castes (Grüter et al. 2012) and replacement behavior (Baudier et al. 2019) have been described only recently.

3 Mathematical analysis

We now analyze the model described in Sect. 2.5 to study how stingless bee colonies regulate colony growth and defense through mechanisms acting on three different timescales. Specifically, we will study how the maturation time of majors and replacement behavior of minors affects the impact of major production on guard availability and colony size.

3.1 Filippov system description

The full model (6) is a Filippov system (Filippov 1988; Meza et al. 2005; da Silveira Costa and Meza 2006; Boukal and Kivan 1999) which can be converted to a generalized form. Let $H(Z) := (G + g) - \theta P$ with vector $Z = (W, G, w, g)^T$, and

$$F_{S_c}(Z) = \begin{pmatrix} \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}\rho - \dfrac{1}{\tau}W \\ \dfrac{1}{\tau}W - \mu_G G \\ \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}(1-\rho) - \mu_w w - \delta w \\ \delta w - \mu_g g, \end{pmatrix}, \tag{7}$$

$$F_{S_b}(Z) = \begin{pmatrix} \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}\rho - \dfrac{1}{\tau}W \\ \dfrac{1}{\tau}W - \mu_G G \\ \Lambda \dfrac{(W+w)^2}{b^2+(W+w)^2}(1-\rho) - \mu_w w \\ -\mu_g g \end{pmatrix}, \tag{8}$$

Then System (6) can be rewritten as the following generalized Filippov system

$$\dot{Z} = \begin{cases} F_{S_c}(Z), & Z \in S_c, \\ F_{S_b}(Z), & Z \in S_b, \end{cases} \tag{9}$$

where $S_c = \left\{Z \in \mathbb{R}_4^+ \mid H(Z) < 0\right\}$, $S_b = \left\{Z \in \mathbb{R}_4^+ \mid H(Z) > 0\right\}$ are two regions divided by the discontinuity manifold

$$\Sigma = \left\{Z \in \mathbb{R}_4^+ \mid H(Z) = 0\right\}.$$

We denote System (6) defined in region S_c as *Crisis Mode with Replacement* and System (6) defined in region S_b as *Non-Crisis Mode without Replacement*. System (9) is piecewise smooth with a discontinuity boundary where each point $Z \in \Sigma$ can be associated with two vectors, $F_{S_c}(Z)$ and $F_{S_b}(Z)$. Orbits of System (9) cross the boundary when the transversal components of $F_{S_c}(Z)$ and $F_{S_b}(Z)$ have the same sign, and in that case they have a discontinuity in their tangent vector. Conversely, when the transversal components have opposite signs, orbits slide on the boundary. Filippov's Convex Method is widely used to describe this sliding motion, and we will implement it in Sect. 3.5. Thus, the state portrait of System (6) consists of the standard state portaits in regions S_c and S_b and the sliding state portait on the discontinuity boundary Σ (Kuznetsov et al. 2003).

Definition 3.1 introduces two types of equilibria relevant to this context (Di Bernardo et al. 2008; Kuznetsov et al. 2003).

Definition 3.1 A point Z^* is called a *regular* equilibrium of System (6) iff

$$F_{S_c}(Z^*) = 0,\, H(Z^*) < 0 \text{ or } F_{S_b}(Z^*) = 0,\, H(Z^*) > 0.$$

Alternatively, a point Z^* is called a *virtual* equilibrium of System (6) iff

$$F_{S_c}(Z^*) = 0, H(Z^*) > 0 \text{ or } F_{S_b}(Z^*) = 0, H(Z^*) < 0.$$

Virtual equilibria are fixed points of F_{S_i}, $i = \{c, b\}$, that lie *outside* S_i. Because System (9) follows F_{S_i} only within S_i, system trajectories cannot not converge to virtual equilibria. Thus, only regular equilibria represent biologically meaningful states of System (6).

In the following sections, we will show that trajectories of System (6) are positive and bounded for positive initial conditions (Sect. 3.2). We will then prove the existence of a trivial extinction equilibrium (Sect. 3.3) and, lastly, we will use the definitions above to analyze the model dynamics in each mode, with and without replacement (Sect. 3.4).

3.2 Positive invariance

Toward validating the biological plausibility of (6), we first prove Theorem 3.1 (in Appendix B.1), which states that none of the population variables become negative over trajectories of the system.

Theorem 3.1 (Basic dynamical properties) *There exists a unique forward solution for System* (6) *starting form each initial condition in domain* $\mathbb{R}^4_+$. *Moreover,* $\mathbb{R}^4_+$ *is positively invariant with respect to System* (6) *and System* (6) *is bounded in* $\mathbb{R}^4_+$. *[Proof in Appendix* B.1*]*

3.3 Stability of the extinction equilibrium

The Full System (6) has a trivial extinction equilibrium at the origin; that is:

$$E^e := (W^*, G^*, w^*, g^*) = (0, 0, 0, 0).$$

The extinction equilibrium is always an equilibrium of both F_{S_c} and F_{S_b}. It may be approached from S_b or S_c, that is, from $G + g > \theta(G + g + W + w)$ or $G + g < \theta(G + g + W + w)$. In fact, a positive neighborhood of the origin contains both points in S_c and S_b. By linearizing F_{S_c} and F_{S_b} around the origin, we can study the stability of E^e in S_c and S_b (i.e., whether a small population would grow and persist or instead decline back to extinction). Theorem 3.2 states that the extinction equilibrium is locally attractive both when approached from S_c and when approached form S_b and also gives sufficient conditions for when E^e is globally stable.

Theorem 3.2 (Extinction equilibrium) *Model* (6) *always has the extinction equilibrium* $E^e := (W^*, G^*, w^*, g^*) = (0, 0, 0, 0)$. *which is always locally stable. Let* $\mu_u = \min\{1/\tau, \mu_w\}$ *and* $\Lambda < 2\mu_u b$. *Then* E^e *is globally stable. [Proof in Appendix* B.2*]*

The condition for global stability of E^e is not dependent upon the initial population size; it only depends on model parameters. In particular, the condition of an upper bound on egg-laying rate ($\Lambda < 2\mu_u b$) is equivalently a condition that the outflow of workers from the non-guarding compartments W and w must be greater than the inflow into these compartments at the half-saturation population b; that is:

$$\frac{\Lambda}{2} < \mu_u b.$$

In words, by definition of half-saturation population b, when there are b non-guarding workers, the eclosion rate that generates new non-guarding workers is $\Lambda/2$. However, each of those workers either dies (at rate μ_w) or transitions to guarding (at rate $1/\tau$) out of non-guarding work at a rate of at least μ_u. Consequently, if the inflow $\Lambda/2$ of new workers into w and W is dominated by the outflow $\mu_u b$ from these groups, then the incipient population will decline over time.

However, even when the conditions for global stability of the extinction equilibrium E^e are not met, E^e is always locally stable by Theorem 3.2. Consequently, if an incipient population is not sufficiently large to escape the basin of attraction of E^e, the population will always go extinct. In fact, in the following sections, we will show that the system exhibits a strong Allee Effect (Stephens and Sutherland 1999); the system has both an upper, stable equilibrium as well as a lower, unstable equilibrium akin to a minimum-population threshold below which population growth is negative and ultimately evolves toward extinction.

3.4 Interior equilibria

Now we analyze the system in search for positive attractors of F_{S_c} and F_{S_b}, that is, demographic equilibria towards which the population evolves when it does not go extinct. Note that F_{S_b} can be studied as a special case of F_{S_c} where the replacement rate is zero ($\delta = 0$). Thus, we start our analysis with the Crisis Mode with Replacement, governed by F_{S_c}, and then draw analogous conclusions for the Non-Crisis Mode without Replacement (F_{S_b}).

In the case when there is an insufficient number of guards (i.e., $H(Z) \leq 0$, or $(G+g)/P \leq \theta$), the minor-worker w-to-g replacement mechanism activates. In this state, the Full Model (6) behaves as the system $\dot{Z} = F_{S_c}$ defined in Eq. (7). Any equilibrium $E^c = (W^*, G^*, w^*, g^*)$ of F_{S_c} must satisfy the following relations:

$$\begin{cases} \mu_G G^* = \dfrac{1}{\tau} W^* \\ \rho(\mu_w + \delta) w^* = \dfrac{1-\rho}{\tau} W^* \\ \mu_g g^* = \delta w^* \end{cases} \tag{10}$$

The relations in System (10) allow us to establish the condition for an equilibrium point to lay in the region of the state space where replacement occurs (S_c), that is, in order to be a *regular equilibrium* of the Full System (6). Define γ_c as the guard ratio $(G^* + g^*)/P^*$ at E^c, that is

$$\gamma_c := \frac{\rho(\mu_w + \delta) + \mu_G \delta (1-\rho)/\mu_g}{\rho(\mu_w + \delta)(1 + \tau \mu_G) + (1-\rho)(1 + \delta/\mu_g)\mu_G}. \tag{11}$$

Then E^c is a regular equilibrium of Full System (6) if and only if $H(E^c) \leq 0$ or, equivalently, $\gamma_c \leq \theta$.

Furthermore, the equilibrium number of major workers W^* must satisfy:

$$\Lambda \frac{\left(W^* + \overbrace{\frac{(1-\rho)W^*}{\tau\rho(\mu_w+\delta)}}^{w^*}\right)^2}{b^2 + \left(W^* + \underbrace{\frac{(1-\rho)W^*}{\tau\rho(\mu_w+\delta)}}_{w^*}\right)^2}\rho = \frac{1}{\tau}W^*.$$

which simplifies to the quadratic equation

$$\Lambda W^* \left(1 + \frac{(1-\rho)}{\tau\rho(\mu_w+\delta)}\right)^2 \rho = \frac{1}{\tau}b^2 + \frac{1}{\tau}(W^*)^2 \left(1 + \frac{(1-\rho)}{\tau\rho(\mu_w+\delta)}\right)^2. \tag{12}$$

The roots of Eq. (12) are the non-trivial, interior equilibria of F_{S_c} (7), whose stability follows from Theorem 3.3.

Theorem 3.3 (Existence and Stability of the crisis interior equilibria) *Define condition*

$$C_c : \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w+\delta}\right) \geq 1. \tag{13}$$

Then the Crisis System defined by F_{S_c} (7) has two interior equilibria, E^c_+ and E^c_-, if and only if C_c holds. Both have the form

$$(W^*, G^*, w^*, g^*) = (W^*, \frac{1}{\tau\mu_G}W^*, \frac{1-\rho}{\rho\tau(\mu_w+\delta)}W^*, \frac{\delta}{\mu_g}\frac{1-\rho}{\rho\tau(\mu_w+\delta)}W^*)$$

where

$$W^* = \frac{1}{2}\left(\Lambda\rho\tau \pm \sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b(\mu_w+\delta)\tau}{\frac{1-\rho}{\rho} + (\mu_w+\delta)\tau}\right)^2}\right) \tag{14}$$

Moreover,

1. *The interior equilibrium E^c_+ is Locally Asymptotically Stable, and E^c_- is unstable.*
2. *Both E^c_+ and E^c_- are* regular *equilibria of the Full System (6) if and only if*

$$\gamma_c < \theta$$

for γ_c defined in Eq. (11).

[Proof in Appendix B.3*]*

In the special case when replacement rate $\delta = 0$, the vector field F_{S_c} (7) reduces to F_{S_b} (8). This allows us to analyze the Non-Crisis Mode without Replacement governed by F_{S_b} (8). Define $C_b := C_c|_{\delta=0}$ and

$$\gamma_b := \gamma_c|_{\delta=0} = \frac{\rho\mu_w}{\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau)}, \tag{15}$$

which can also be written as

$$\gamma_b = \overbrace{\rho}^{\substack{\text{fraction}\\ \text{of majors}}} \frac{\overbrace{1/\mu_G}^{\substack{\text{guarding time}\\ \text{for majors}}}}{\underbrace{(1-\rho)/\mu_w + \rho(1/\mu_G + \tau)}_{\text{average bee lifespan}}}$$

Now we provide conditions for the existence and stability of interior equilibria in the Non-Crisis System F_{S_b} (8).

Theorem 3.4 (Existence and stability of the non-crisis interior equilibria)
Define condition

$$C_b : \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w}\right) \geq 1. \tag{16}$$

Then the Non-Crisis System defined by F_{S_b} (8) *has two interior equilibria $E^b_+ = E^c_+|_{\delta=0}$ and $E^b_- = E^c_-|_{\delta=0}$ if and only if C_b holds.*
Moreover,

1. *The interior equilibrium E^b_+ is Locally Asymptotically Stable, while E^b_- is unstable.*
2. *Both E^b_+ and E^b_- are* regular *equilibria of the Full System* (6) *if and only if*

$$\gamma_b > \theta$$

for γ_b defined in Eq. (15).

[Proof in Appendix B.4*]*

As with Theorem 3.3 in the crisis mode, Theorem 3.4 (in the non-crisis mode) provides a sufficient condition for the existence of an interior regular attractor E^b_+ when the proportion of guards $(G+g)/P$ in the colony is above the replacement threshold θ (see Table 3 for a summary of the results from Theorems 3.3–3.4).

By Theorem 3.2, the extinction equilibrium E^e is always a local attractor of the full System (6). Theorem 3.4 provides a sufficient condition when the no-replacement Model (8) has a second attractor, E^b_+, separated from E^e by an unstable node E^b_-.

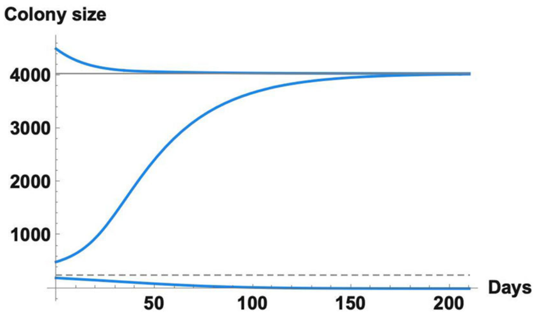

Fig. 2 The system exhibits an Allee Effect. *Thick blue* lines represent trajectories with the same parameter values but different initial conditions. If the initial population is below the unstable equilibrium (*dashed horizontal line*) the population collapses. Otherwise, it establishes at the stable equilibrium (*solid horizontal line*). Parameter values: $\Lambda = 154$, $b = 1000$, $\rho = 0.06$, $\tau = 20$, $\mu_G = 1/5.4$, $\mu_g = 1/3$, $\mu_w = 1/28$, $\theta = 0.01$, $\delta = 4$, $W(0) = G(0) = g(0) = 0.w(0) = 200, 500, 4500$, respectively

Figure 2 illustrates such an example where the colony can survive if: (1) Condition C_b is satisfied, and (2) the initial population is above certain threshold related to E_-^b (see figure caption for parameter values).

From condition C_b of Theorem 3.4 for non-crisis System (8) without replacement, the non-trivial stable equilibrium E_+^b can only exist when $\Lambda/2 \geq b/(\rho\tau + (1-\rho)/\mu_w)$; otherwise, the colony will collapse. That is, when there are b workers, the spontaneous regeneration rate due to eclosion ($\Lambda/2$) must be at least the average rate of attrition due to guard maturation or worker death ($b/(\rho\tau + (1-\rho)/\mu_w)$) in order to allow for the colony to grow to the non-trivial, stable equilibrium E_+^b from a sufficiently large initial population size is sufficiently large. Furthermore, because of the ever-present local stability of the extinction equilibrium E^e, even when C_b is satisfied, if the population suddenly drops below a critical number (related to E_-^b), the colony can still go extinct.

Assuming that C_b is satisfied and the initial population size is sufficiently large, the colony population will arrive at the stable equilibrium E_+^b, and the total population size will be:

$$W^* \left(1 + \frac{1-\rho}{\rho\mu_w\tau} + \frac{1}{\tau\mu_G}\right). \tag{17}$$

At E_+^c and E_+^b, the value of W^* given in (14) (with $\delta = 0$ for E_+^b) is increasing with Λ, the egg-laying rate, and decreasing with b, the non-guarding population for eclosion half-saturation. This means that the higher the queen's egg-laying rate and the smaller the number of non-guarding workers needed to ensure brood survival to eclosion, the larger a mature colony can become. However, if the population suddenly falls, the colony may move into the basin of attraction of extinction equilibrium E^e (discussed above); alternatively, the colony may move into crisis, guard-replacement mode if the guard ratio falls below threshold θ.

Remark 3.1 Note that, when $\delta > 0$, condition C_c automatically satisfies condition C_b because

$$\underbrace{\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta}}_{*} \leq \underbrace{\rho\tau + (1-\rho)\frac{1}{\mu_w}}_{\dagger}.$$

That is, the average time for a non-guarding worker to transition to guarding in the crisis mode ($*$ above) is less than in the non-crisis mode ($\dagger$ above). Consequently, by Theorem 3.3, persistence of the colony in crisis mode requires a faster egg-laying rate Λ or a reduction in the number of non-guarding workers necessary to ensure a high probability of development of eggs to eclosion.

3.4.1 Virtual and regular equilibria

As noted in Theorems 3.3 and 3.4, the fraction $(G+g)/P$ of guards in the colony determines whether the equilibria are virtual or regular. Because only regular equilibria can be approached by the system's trajectories, a complete analysis of the system's dynamics requires further characterizing the guard ratios γ_c and γ_b.

Theorem 3.5 (Characterization of the equilibrium guard ratios)
Consider the ratios

$$\gamma_b = \frac{\rho\mu_w}{\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau)}$$

defined in Eq. (15) *and*

$$\gamma_c = \frac{\rho\left(\mu_w + \delta\right) + \mu_G(1-\rho)\frac{\delta}{\mu_g}}{\mu_G(1-\rho)\left(\frac{\delta}{\mu_g} + 1\right) + \rho\left(\mu_w + \delta\right)\left(1 + \tau\mu_G\right)}.$$

defined in Eq. (11). Then the following are true:

1. *The ratio γ_b is monotonically increasing with respect to ρ.*
2. *The ratio γ_c is monotonically decreasing with respect to ρ if $\mu_g \leq \delta\tau\mu_G$ while it is monotonically increasing otherwise.*
3. *Moreover, we always have*

$$\gamma_b \leq \gamma_c,$$

which implies that two positive, stable, regular equilibria cannot coexist in the Full System (6).

[Proof in Appendix B.5*]*

The fact that the base ratio of guards γ_b is increasing with respect to ρ is natural given that, without replacement, all guards are majors and therefore increasing ρ directly increases the proportion of guards in the colony. However, in the crisis system, the effect of increasing ρ on the guard proportion depends on the parameter values because the guard population is a mix of both minors and majors. The condition $\mu_g \leq \delta\tau\mu_G$ or, equivalently,

$$\frac{1/\tau}{\mu_G} \leq \frac{\delta}{\mu_g}$$

can be interpreted as follows: majors mature at a rate $1/\tau$ into the guard group G and perform defense tasks for an average time $1/\mu_G$. Analogously, minors are recruited into the guard group g at a rate δ and perform the task for $1/\mu_g$ time units. If $(1/\tau)/\mu_G \leq \delta/\mu_g$, this means that there are more recruitment events in a guard's lifespan for minors than for majors, and therefore increasing the proportion of majors ρ ultimately decreases the total proportion of guards in the colony (γ_c is decreasing with respect to ρ). Conversely, if $(1/\tau)/\mu_G > \delta/\mu_g$, then the majors have more recruitment events per guard lifespan, and thus increasing major production (ρ) increases the fraction of guards in the colony γ_c.

3.5 Dynamics on the switching boundary Σ

We have studied the dynamics of the crisis mode F_{S_c} (7), where minors adopt guarding tasks in response to a lack of soldiers, and the non-crisis mode F_{S_b} (8) (without replacement by minors). Each of these systems has two interior equilibrium points, one stable and one unstable, which may both be virtual or regular according to the value of the replacement threshold θ. It remains to analyze the system behavior on the switching boundary Σ. Trajectories of the Full System (6) may remain in one of the regions S_c and S_b, cross the switching surface Σ or slide along it. In order to investigate the crossing and sliding dynamics, we first determine the existence of a *crossing set* and a *sliding set* on Σ (Filippov 1988; da Silveira Costa and Meza 2006; Boukal and Kivan 1999; Tang et al. 2012a, b; Xiao et al. 2013).

Let

$$\sigma(Z) := \langle H_z(Z), F_{S_c}(Z)\rangle\langle H_z(Z), F_{S_b}(Z)\rangle, \tag{18}$$

where $\langle\cdot\rangle$ denotes the standard scalar product and $H_z(Z)$ is the non-vanishing gradient of smooth function H on Σ. Define the *crossing set* $\Sigma_C \subset \Sigma$ as

$$\Sigma_C = \{Z \in \Sigma \mid \sigma(Z) > 0\},$$

and the *sliding set* $\Sigma_S \subset \Sigma$ as

$$\Sigma_S = \{Z \in \Sigma \mid \sigma(Z) \leq 0\},$$

where $\Sigma_S = \Sigma \setminus \Sigma_C$. For System (6), we can obtain that

$$\begin{aligned}\sigma(Z) = \theta^2 \left(\tfrac{\Lambda(W+w)^2\rho}{(b^2+(W+w)^2)} - W/\tau\right)^2 + (1-\theta)^2(W/\tau - \mu_G G)^2 \\ +\theta^2\left(\tfrac{\Lambda(W+w)^2(1-\rho)}{(b^2+(W+w)^2)} - \mu_w w\right)^2 + (1-\theta)^2(g\mu_g)^2 \\ +\delta w\left(-\theta^2\left(\tfrac{\Lambda(W+w)^2(1-\rho)}{(b^2+(W+w)^2)} - \mu_w w\right) - (1-\theta)^2 g\mu_g\right)\end{aligned}$$

for all $Z \in \Sigma$, that is, for

$$\left\{Z \in \mathbb{R}_4^+ \mid H(Z) = 0\right\} = \left\{Z \in \mathbb{R}_4^+ \mid (G+g) - \theta(G+g+W+w) = 0\right\}.$$

The boundary Σ may also contain particular points, called "pseudoequilibria" (Di Bernardo et al. 2008), that act as equilibria within the sliding set Σ_S. In order to characterize them, we follow Filippov's Convex Method to define sliding motions on Σ_S as the solutions to the continuous ODE $\dot{Z} = F_0(Z)$ on Σ_S, where $Z \in \Sigma_S$ and $F_0(Z)$ is a convex combination of $F_{S_c}(Z)$ and $F_{S_b}(Z)$ tangent to Σ at Z (Kuznetsov et al. 2003). The vector field $F_0(Z)$ thus has the form

$$F_0(Z) = \lambda(Z)F_{S_c}(Z) + (1-\lambda(Z))F_{S_b}(Z) \tag{19}$$

where $\lambda : \Sigma \to [0, 1]$ is a function of Z defined so that $F_0(Z)$ is tangent to Σ_S, that is,

$$\lambda(Z) = \frac{\langle H_z(Z), F_{S_b}(Z)\rangle}{\langle H_z(Z), F_{S_b}(Z) - F_{S_c}(Z)\rangle}$$

Definition 3.2 A point $Z \in \Sigma_S$ is a *pseudoequilibrium* if $F_0(Z) = 0$.

The vector field $F_0(Z)$ can be expressed as below:

$$F_0(Z) = \begin{pmatrix} \Lambda\dfrac{(W+w)^2}{b^2+(W+w)^2}\rho - \dfrac{1}{\tau}W \\ \dfrac{1}{\tau}W - \mu_G G \\ \Lambda\dfrac{(W+w)^2}{b^2+(W+w)^2}(1-\rho) - \mu_w w - (1-\lambda)\delta w \\ (1-\lambda)\delta w - \mu_g g. \end{pmatrix} \tag{20}$$

It follows that a pseudoequilibrium $E^p = (W^*, G^*, g^*, w^*)$ must satisfy the relations

$$\begin{cases} G^* = \dfrac{1}{\tau\mu_G} W^*, \\ g^* = \dfrac{1-\rho}{\rho\tau(\mu_w + \delta(1-\lambda^*))} W^*, \text{ and} \\ w^* = \dfrac{\delta(1-\lambda^*)}{\mu_g} \dfrac{1-\rho}{\rho\tau(\mu_w + \delta(1-\lambda^*))} W^*, \end{cases} \tag{21}$$

where $\lambda^* = \lambda(E^p)$. From this relations, we can find the ratio of guards at pseudoequilibria. Define γ_p as the ratio $\frac{G^*+g^*}{P^*}$ for G^*, g^* and $P^* = W^* + w^* + G^* + g^*$ as in Eq. (21) above, that is,

$$\gamma_p := \frac{\rho(\mu_w + \delta(1-\lambda^*)) + \mu_G\delta(1-\lambda^*)(1-\rho)/\mu_g}{\rho(\mu_w + \delta(1-\lambda^*))(1+\tau\mu_G) + (1-\rho)(1+\delta(1-\lambda^*)/\mu_g)\mu_G}. \tag{22}$$

Note that, to be contained in Σ, the pseudoequilibrium must satisfy the condition

$$\gamma_p = \theta.$$

From this equality we can find the value of $\lambda^* = \lambda(E^p)$ as

$$\lambda^* = \frac{\mu_G\mu_g\theta(\delta\tau\rho + 1 - \rho + \mu_w\rho\tau) - (1-\theta)(\delta(\mu_G(1-\rho) + \mu_g\rho) + \mu_g\mu_w\rho)}{\delta(\mu_G\mu_g\rho\tau\theta - (1-\theta)(\mu_G(1-\rho) + \mu_g\rho))} \tag{23}$$

These relations yield the following Theorem 3.6, which establishes the existence of two pseudoequilibria on Σ when the Full System (6) has no regular equilibria.

Theorem 3.6 (Existence of pseudoequilibria) *Define condition*

$$C_p : \frac{\Lambda}{2b} \frac{(1-\theta)(\mu_g\rho + \mu_G(1-\rho) + \mu_G\mu_w\tau\rho)}{\mu_G(\theta\mu_g + \mu_w(1-\theta))} \geq 1. \tag{24}$$

Then System (6) has two pseudoequilibria E^p_+ and E^p_- on the switching surface Σ if and only if C_p holds and $\gamma_b \leq \theta < \gamma_c$. The pseudoequilibria have the form

$$(W^*, G^*, g^*, w^*) = \left(W^*, \frac{1}{\tau\mu_G} W^*, \frac{(1-\theta)(\mu_G(1-\rho) - \mu_g\rho) + \theta\mu_G\mu_g\rho\tau}{\mu_G\rho\tau(\theta\mu_g + \mu_w(1-\theta))} W^*, \right.$$
$$\left. \frac{-\mu_w\rho + \theta(\mu_G(1-\rho) + \mu_w\rho(1+\mu_G\tau))}{\mu_G\rho\tau(\theta\mu_g + \mu_w(1-\theta))} W^* \right)$$

with

$$W^*|_{E^p_\pm} = \frac{1}{2}\rho\tau\left(\Lambda \pm \sqrt{\Lambda^2 - \left(\frac{2b\mu_G(\theta\mu_g + (1-\theta)\mu_w)}{(1-\theta)(\mu_G(1-\rho) + \mu_g\rho + \mu_G\mu_w\rho\tau)}\right)^2}\right)$$

[Proof in Appendix B.6]

The pseudoequilibrium existence condition C_p can be derived from the crisis existence condition C_c (3.4) by substituting the value of δ for $\delta(1-\lambda^*)$. Indeed, C_p can be written as

$$\frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta(1-\lambda^*)}\right) \geq 1$$

which, like C_c and C_b, sets a minimum egg-laying rate Λ to ensure that there are enough individuals contributing to egg eclosion (W and w compartments) despite major maturation and worker death (see discussion of Theorems 3.3–3.4). For $\gamma_b \leq \theta \leq \gamma_c$, condition C_p is met automatically when

$$C_c : \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta}\right) \geq 1$$

holds, because λ^* takes values between 0 and 1 and thus

$$\frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta(1-\lambda^*)}\right) \geq \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta}\right).$$

Similarly, condition C_p implies condition C_b from Theorem 3.4 because

$$\frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w}\right) \geq \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta(1-\lambda^*)}\right).$$

Therefore, we have

$$C_c \implies C_p \implies C_b.$$

4 Interacting defense mechanisms

Now we study the interaction between the three mechanisms determining guard allocation, namely the major production regulated by ρ, the age-based polyethism regulated by the maturation time τ, and the replacement by minors that occurs when the guard proportion drops below θ. In particular, we are interested in determining

1. The effect of these three parameters on colony size and task allocation
2. The parameter combinations keep the system out of crisis, that is, governed by the Non-Crisis System (8).

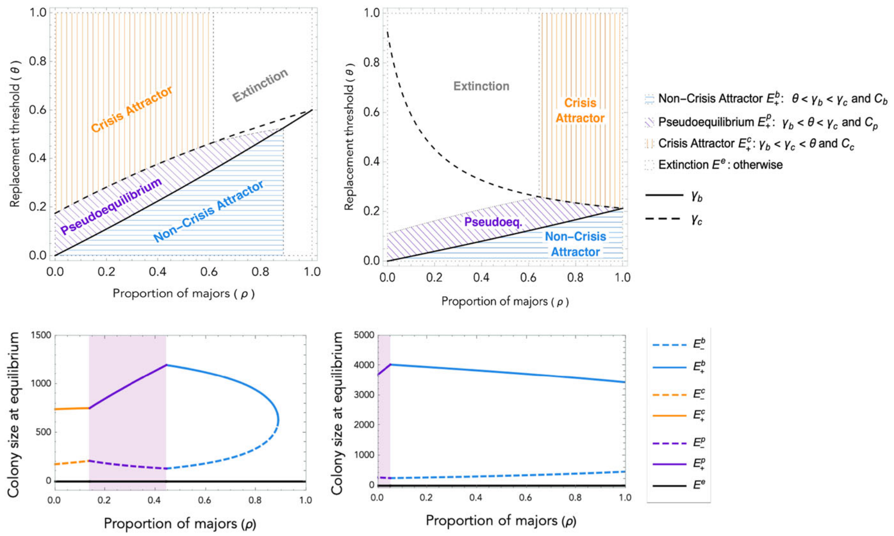

Fig. 3 Model behavior with respect to ρ and θ. *Top panels*: In different regions of the ρ vs θ space, the system may have a crisis attractor, a pseudoequilibrium, a non-crisis attractor (with no minor guards), or only an extinction equilibrium. The crisis and non-crisis guard ratios, γ_c and γ_b, separate the regions. By Theorem 3.5, γ_c might be increasing (*top left panel*) or decreasing (top right panel). *Bottom panels*: Below each of the *top panels*, the corresponding bifurcation diagram for colony size with respect to ρ is shown for a fixed value of θ. In the *bottom left* panel, $\theta = 0.25$ and as ρ increases, the system transitions from crisis to pseudoequilibrium, to non-crisis, where the population decreases until it suddenly drops to extinction. In the *bottom right* panel, $\theta = 0.01$ and the system transitions from pseudoequilibrium, with an increasing population, to non-crisis, where the colony size decreases but does not reach an extinction state. Parameter values: (*left panels*) $\Lambda = 50, b = 300, \tau = 10, \mu_G = 1/15, \mu_g = 1/7, \mu_w = 1/28, \delta = 0.03$; (*right panels*) $\Lambda = 154, b = 1000, \tau = 20, \mu_G = 1/5.4, \mu_g = 1/3, \mu_w = 1/28, \delta = 4$. The parameters in the *right panels* are approximated values for real *T. angustula* colonies (see Table 2 and Appendix C)

4.1 Behavioral plasticity

The parameter θ is the threshold fraction of guards in the colony below which minors will adopt guarding tasks. It can be interpreted as the colony demand for guards (the higher θ, the higher the demand), or as the degree of plasticity of minors (the higher θ, the more likely minors are to switch to guarding tasks). Because, as noted in Theorem 3.5, the equilibrium ratio of guards γ_c in the crisis regime (with replacement) is always greater than the corresponding ratio γ_b in the non-crisis regime (without replacement), there are only three possible alternatives for the threshold ratio θ and they determine the system's behavior, as illustrated in Fig. 3.

1. $\gamma_b \leq \gamma_c \leq \theta$
   By Theorem 3.6, there are no pseudoequilibria. Then we have the following cases based on condition C_c:

   (a) C_c holds: **Crisis** (E_+^c)
   - Condition C_c implies the existence of the Crisis Attractor E_+^c, which is a regular equilibrium of the Full System (6) because $\gamma_c \leq \theta$ (Theorem 3.3);

- Condition $C_c \implies C_b$, so the Non-Crisis Attractor E_+^b exists, but as a virtual equilibrium of the Full System (6) because $\gamma_b \leq \theta$ (Theorem 3.4);

Thus, because of the Allee Effect, trajectories either converge to E_+^c or collapse to E^e if the population drops below a critical size.

(b) C_c does not hold ($\neg C_c$): **Extinction** (E^e)
- There is no Crisis Attractor E_+^c;
- The Non-Crisis Attractor E_+^b exists as a virtual equilibrium of the Full System (6) if C_b holds and does not exist otherwise;

It follows that the only attractor is the extinction equilibrium E^e.

The ratio of guards at equilibrium, even with replacement, is less than θ. Then the model converges to the Crisis Attractor E_+^c (like the orange "Crisis" trajectories in Fig. 4) or extinction E^e. Biologically, this would represent a situation where minors keep transitioning to guarding even when all the major and replacement guards that the colony can produce are defending the nest. This is typically not observed in real colonies, where the guarding population consists mainly of major bees.

2. $\theta < \gamma_b \leq \gamma_c$
By Theorem 3.6, there are no pseudoequilibria. Then we have the following cases based on condition C_b:

(a) C_b holds: **Non-crisis** (E_+^b)
- Condition C_b implies the existence of the Non-Crisis Attractor E_+^b, which is a regular equilibrium of the Full System (6) because $\theta \leq \gamma_b$ (Theorem 3.4);
- The Crisis Attractor E_+^c exists as a virtual equilibrium of the Full System (6) if C_c holds and does not exist otherwise;

Thus, trajectories either converge to E_+^b or collapse to E^e if the population drops below the Allee threshold.

(b) C_b does not hold ($\neg C_b$): **Extinction** (E^e)
- The Non-Crisis Attractor E_+^b does not exist;
- Because $\neg C_b \implies \neg C_c$, there is no Crisis Attractor E_+^c;

It follows that the only attractor is the extinction equilibrium E^e and the population collapses.

The ratio of guards at equilibrium is greater than θ, even without replacement. Then the model converges to the Non-Crisis Attractor E_+^b (like the blue "Non-crisis" trajectories in Fig. 4) or extinction E^e. This is the typical situation for colonies in the field: enough guards are produced such that minor replacement is not required. Replacement would be activated upon guard removal, and would accelerate the recovery of the base guard proportion in the colony, γ_b. However, even without replacement, the original proportion of guards would be eventually reached.

3. $\gamma_b \leq \theta < \gamma_c$
The Crisis and Non-crisis Attractors E_+^c and E_+^b exist as virtual equilibria of the Full System (6) if their respective existence conditions C_c and C_b are met, and do not exist otherwise. Then we have the following cases based on condition C_p:

(a) C_p: **Regulating** (E_+^p)
- Condition C_p implies the existence of pseudoequilibrium $E_+^p \in \Sigma_s$ because $\gamma_b \leq \theta < \gamma_c$ (Theorem 3.6);

Thus, trajectories either converge to the pseudoequilibrium E_+^p or collapse to E^e if the population drops below a critical size.

(b) C_p does not hold $(\neg C_p)$: **Extinction** (E^e)
- No pseudoequilibria exist.

It follows that the only attractor is the extinction equilibrium E^e and the population collapses.

This would represent a colony where the natural production of major guards is too low, which triggers replacement by minors, but the mixed guarding population resulting from replacement is large enough to prevent further replacements. If the existence condition is not satisfied, the population collapses to extinction, like in the previous cases (see gray "Extinction" trajectories in Fig. 4). However, convergence to a pseudoequilibrium in this case implies a persistent population of minor guards (see purple "Pseudoequilibrium" trajectories in Fig. 4), which is not typically observed in real colonies.

The simulations in Fig. 4 illustrate system trajectories converging to the "Crisis", "Pseudoequilibrium", "Non-crisis", and "Extinction" attractors. The parameter values are typical values for *Tetragonisca angustula* stingless bee colonies (see Table 2), the same values used in the top right panel of Fig. 3. Note how there are no minor guards at equilibrium in the "Non-crisis" attractor, but there is a persistent population of them for the "Crisis" and "Pseudoequilibrium" attractors. This is not efficient because minor guards have a lower performance at defense tasks and thus are meant to be a back-up for exceptional situations where major guards are missing. In fact, although real colonies in the field present a mixed guard population after guard removal (Baudier et al. 2019), they usually have major bees defending the nest.

The value of θ must be less than γ_b in order for the system to operate in non-crisis mode. The maximum value of γ_b, the fraction of guards without replacement, is at $\rho = 1$, with an all-major colony. Thus, θ must be less than $\gamma_b|_{\rho=1} = 1/(1 + \mu_G\tau) = (1\mu_G)/(\tau + 1/\mu_G)$, which is the fraction of their lives that majors spend as guards. If θ is greater than this proportion, then the colony is guaranteed to either go extinct or maintain a persistent population of minor inefficiently allocated to guarding tasks. Moreover, if the colony is in a crisis state (orange regions in Fig. 3), decreasing θ will always lead the system to a non-crisis mode (blue regions in Fig. 3) but increasing θ will not force the colony into extinction. However, increasing θ from a non-crisis state may lead to either crisis mode, with a persistent population of minor guards, or directly to extinction, without ever transitioning to a crisis mode. That is, a colony that can sustain itself in the crisis attractor, with a stable population of minors dedicated to guarding, will not collapse because of demographic factors if the demand for guards increases or minors become exceedingly plastic (meaning that they would adopt guarding tasks even if the whole population was already guarding). This is because, in crisis, the colony is simply unable to meet its demand for guards $(\gamma_c < \theta)$, so increasing this demand makes no difference for task allocating or colony size (the colony is already

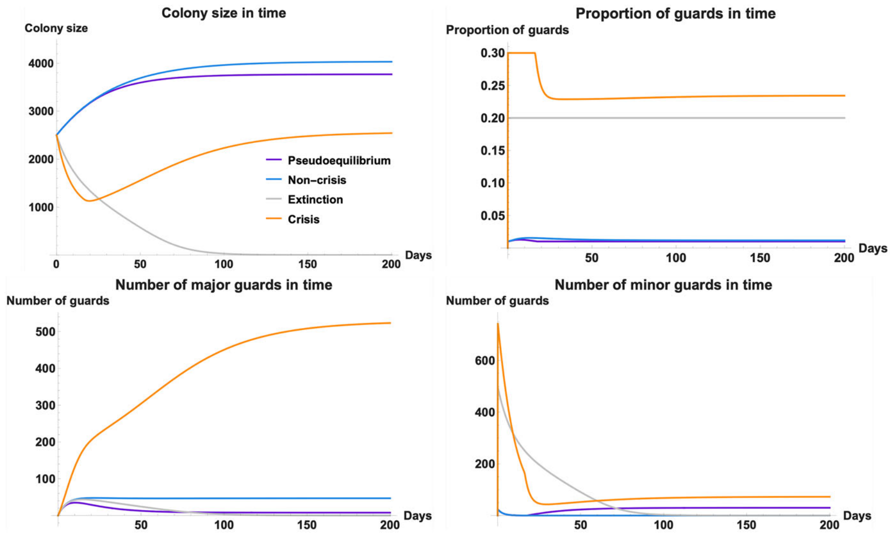

Fig. 4 System trajectories illustrating the "Non-crisis"($\rho = 0.06, \theta = 0.01$), "Pseudoequilibrium" ($\rho = 0.01, \theta = 0.01$) and "Crisis" ($\rho = 0.8, \theta = 0.3$) and "Extinction"($\rho = 0.06, \theta = 0.2$) attractors. Each trajectory has a different combination of ρ and θ corresponding to one of the four regions in the 2-parameter bifurcation in the *top right panel* of Fig. 3. Note the persistent population of minor (replacement) guards in the "Crisis" and "Regulating" cases. ($\Lambda = 154, b = 1000, \tau = 20, \mu_G = 1/5.4, \mu_g = 1/3, \mu_w = 1/28, \delta = 4$)

producing all the major and replacement guards it can). However, it may become exposed to higher environmental risks, like robbery and attacks.

4.2 Morphological specialization

The fraction ρ represents the proportion of eggs that eclose into majors, which mature into soldiers. As seen in Fig. 3, for sufficiently small values of θ, increasing ρ can keep the system in non-crisis mode, without replacement guards. However, as shown in the bottom panels of Fig. 3, increasing ρ may also reduce colony size and even lead to colony extinction. The effect of modifying the production of majors ρ depends on the value of θ and whether γ_c is increasing or decreasing. There are six possible cases illustrated in Fig. 5. Define ρ_C as the value of ρ, if any, such that $\gamma_c|_{\rho=\rho_C} = \theta$. Correspondingly, define ρ_B as the value of ρ, if any, such that $\gamma_b|_{\rho=\rho_B} = \theta$.

1. Increasing γ_c ($\gamma_b|_{\rho=0} < \gamma_c|_{\rho=0} < \gamma_c|_{\rho=1} = \gamma_b|_{\rho=1}$)

   (a) High value of θ:

$$\theta \geq \gamma_c|_{\rho=1} = \frac{1}{1 + \mu_G \tau}$$

   The system is in the "Crisis" state for all values of ρ, but increasing ρ increases the proportion of guards at equilibrium.

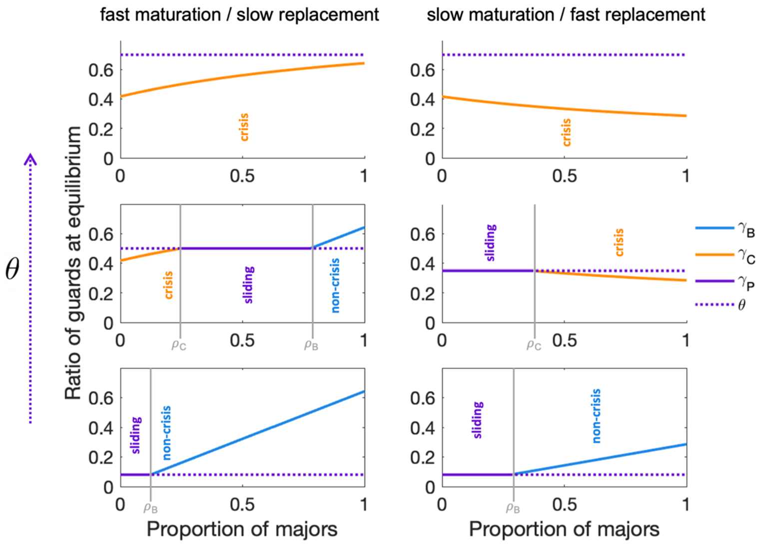

Fig. 5 Proportion of guards at equilibrium for different combinations of ρ and θ. The system converges to the "Crisis" mode (7), "Sliding" mode (20) or "Non-Crisis" mode (8) as the proportion of majors ρ changes. ρ_C and ρ_B are the values of ρ, if any, such that $\gamma_c|_{\rho=\rho_C} = \theta$ and $\gamma_b|_{\rho=\rho_B} = \theta$, respectively. In all panels, $b = 100$, $\mu_g = 1/10$, $\mu_w = 1/28$, $\Lambda = 100$, $\delta = 1/14$. In the *left* column γ_c is increasing, with $\tau = 10$ and $\mu_G = 1/(28 - \tau)$ so that majors and minors have an average lifespan of 28 days, and the θ values are 0.7 in the first row, 0.5 in the second and 0.08 in the third. The *right* column shows decreasing γ_c, with $\tau = 20$, $\mu_G = 1/(28 - \tau)$. The θ values are 0.7 in the first row, 0.35 in the second and 0.08 in the third

(b) Intermediate value of θ:

$$\frac{\delta}{\delta + \mu_g} = \gamma_c|_{\rho=0} < \theta < \gamma_b|_{\rho=1} = \frac{1}{1 + \mu_G \tau}$$

As ρ increases, the system transitions from the "Crisis" state ($0 < \rho < \rho_C$) to the "Regulating" state ($\rho_C < \rho < \rho_B$) and the "Base" state ($\rho > \rho_B$). In this case, increasing the production of majors is a long-term solution for keeping the system out of crisis.

(c) Low value of θ:

$$0 < \theta < \gamma_c|_{\rho=0}$$

As ρ increases, the system transitions from the "Regulating" state ($0 < \rho < \rho_B$) to the "Base" state ($\rho_B < \rho < 1$). In this case, increasing the production of majors is a long-term solution for **increasing** the guard proportion.

2. Decreasing γ_c ($\gamma_b|_{\rho=0} < \gamma_c|_{\rho=1} = \gamma_b|_{\rho=1} < \gamma_c|_{\rho=0}$)

(a) High value of θ:

$$\theta \geq \gamma_c|_{\rho=1}$$

The system is in the "Crisis" state for all values of ρ, but increasing the proportion of majors ρ **decreases** the proportion of guards in the long term.

(b) Intermediate value of θ:

$$\gamma_c|_{\rho=1} < \theta < \gamma_c|_{\rho=0}$$

As ρ increases, the system transitions from the "Regulating" state ($0 \leq \rho \leq \rho_C$) to the "Crisis" state ($\rho_C < \rho \leq 1$). In this case, increasing the production of majors **decreases** the guard proportion in the long term.

(c) Low value of θ:

$$0 < \theta < \gamma_c|_{\rho=1}$$

As ρ increases, the system transitions from the "Regulating" state ($0 \leq \rho < \rho_B$) to the "Base" state ($\rho_B \leq \rho \leq 1$). In this case, increasing the production of majors is a long-term solution for **increasing** the guard proportion.

These cases reveal the situations where the colony can function normally, in the "Non-Crisis" regime, and where an increment in the major production is actually beneficial for colony defense. For example, if major and minor guards live equally long on average and replacement happens at a faster rate than maturation (decreasing γ_c), then increasing the production of majors enhances colony defense only if the proportion of guards can drop very low without triggering replacement (case 2c). In fact, cases 2a-2c show that the minimum proportion of guards required to defend the colony without replacement (θ) must be less than $\gamma_b|_{\rho=1}$, that is the equilibrium proportion of guards if all bees are majors. If θ is greater than this value, then increasing the production of majors ρ actually decreases the fraction the population dedicated to guarding.

The colony size bifurcation diagrams in the *left and right bottom panels* of Fig. 3 are examples of cases 1b and 2c, respectively. They have the same parameters as the *top panels* directly above them, and shows colony size with respect to ρ for a fixed θ value of 0.25 (left) and 0.01 (right). While the parameters on the *right panels* are biologically realistic for *T. angustula* stingless be colonies, the parameters on the *left panels* were chosen to produce a representative diagram with the four possible dynamical outcomes and are not based on these reference values. They would represent an analogous system (like a different species or an artificial system) where γ_c is increasing, so $\mu_g > \delta\tau\mu_G$ by Theorem 3.5. Thus, the ratio between the death rates of minor and major guards must be greater than the ratio between replacement and maturation rates. In this *bottom left panel* of Fig. 3, corresponding to case 1b, ρ_C and ρ_B are well defined. For ρ values below ρ_C, the colony size is dictated by the crisis model (7) because the parameters satisfy condition (11) but not (15). Similarly, for ρ values above ρ_B the system follows the non-crisis model (8) because only condition (15) holds. However, higher ρ values violate the Existence Condition (3.4) for the Base Model, so the only stable equilibrium is extinction. However, the fraction of guards in the colony increases with ρ in case 1b. Thus, increasing the production of majors in this scenario may reinforce colony defense by increasing the fraction of the population allocated to guarding, but may also reduce the colony size or even cause extinction.

On the other hand, the *bottom right panel* of Fig. 3, which falls under case 2c, represents a realistic situation for stingless bee colonies. Because the death rates for major and minor guards are similar and replacement happens much faster than maturation, we have $\mu_g \leq \delta\tau\mu_G$ and γ_c is increasing by Theorem 3.5. Thus, the system cannot transition between crisis and non-crisis states by changing the production of majors. As seen in the *top right panel* of Fig. 3, for high values of θ, the colony will go extinct for low ρ values and persist in a crisis state for high enough ρ values. However, for the small θ values that presumably exist in real colonies, the system transitions from a pseudoequilibrium to the non-crisis attractor as ρ increases, with the population size peaking at ρ_B, where the transition happens. This might suggest why such a small proportion of majors exists in real colonies: the production is enough to supply the demand indicated by θ but not much greater because this would decrease colony size.

4.3 Maturation time τ

We have mentioned the role of the maturation rate in determining whether the fraction of guards at the crisis attractor, γ_c, is increasing or decreasing with respect to ρ. In real stingless-bee colonies, the maturation rate that is considerably slower than the replacement rate, δ, and so γ_c is increasing like in the right panels of Figs. 3 and 5. However, the maturation rate also has an important role in defining the existence conditions C_b, C_p and C_c. As seen in Fig. 6, decreasing the maturation time constant τ also reduces the region in the θ–ρ space where the colony can survive. Thus, a colony where majors mature faster will be able to sustain less majors because high values of ρ lead to extinction. Also note that the colony size peaks at a smaller value of ρ for faster maturation times. If colonies regulate their major production to be near this peak, then this result suggests that colonies where the transition of majors to guarding happens faster are expected to produce a lower number of majors.

Figure 7 shows the equilibrium task allocation as ρ varies for the same values of τ (10 days and 20 days). For $\tau = 20$, which is a biologically realistic value, increasing ρ produces an increment in the non-guarding and guarding major populations (W and G, respectively, represented by darker lines). There are no minor guards g in the non-crisis case, which constitutes most ρ values (except for the shaded areas, corresponding to a pseudoequilibrium). While there are still no minor guards for $\tau = 10$, the major population does not simply increase with ρ. In fact, both guarding and non-guarding populations peak at a high value of ρ and then decrease before collapsing drastically to extinction. Thus, for very fast maturation rates, increasing the production of majors may, perhaps counterintuitively, decrease the number of guards in the colony.

5 Discussion

In this work, we developed a framework for modeling task allocation for collective defense motivated by the stingless bee *T. angustula*. We studied morphological specialization, age-based polyethism and behavioral plasticity as mechanisms regulating

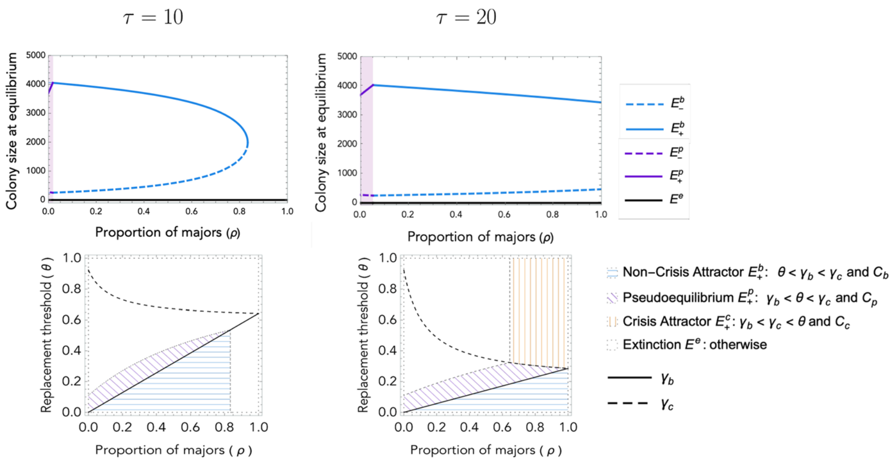

Fig. 6 Effect of τ on colony size bifurcation with respect to ρ (*top*) and system sybamics with respect to ρ and θ (*bottom*). Parameter values: $\Lambda = 154$, $b = 1000$, $\tau = 20$, $\mu_G = 1/5.4$, $\mu_g = 1/3$, $\mu_w = 1/28$, $\delta = 4$. In *top panels*, $\theta = 0.01$

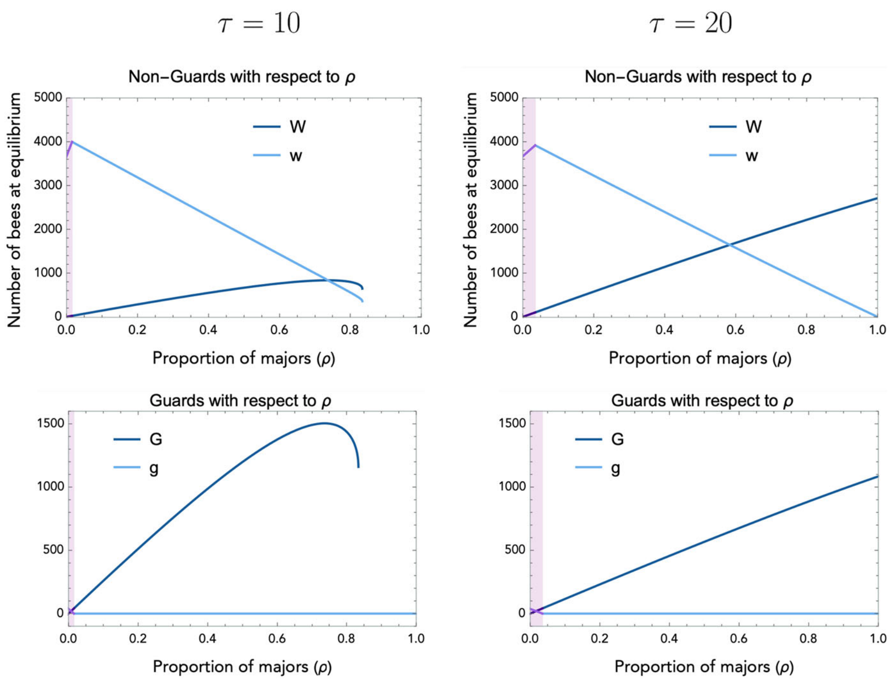

Fig. 7 Bifurcation diagram for all population groups with respect to ρ for $\tau = 10$ and $\tau = 20$. Only stable interior attractors are shown, although extinction is always a stable equilibrium

group defense at different timescales and interacting to maintain colony growth while responding efficiently to crisis situations.

5.1 Colony growth and survival

Our analysis provides basic conditions for colony survival: a non-guarding worker must spend enough time rearing brood to ensure that at least one new adult is recruited before she dies or switches to a guarding task (see Theorem 3.2). Even if this condition is satisfied, the colony may still inevitably collapse if it reaches low numbers, which is known as strong Allee Effect (see Theorems 3.2, 3.4) (Stephens and Sutherland 1999). The need for collective brood care in hives is known to induce Allee effects (Dennis and Kemp 2016). This is the case with stingless-bee colonies, which must assemble swarms of a minimum size in order to produce offspring colonies (Van Veen and Sommeijer 2000). The eclosion term in Eq. (1) has been used to incorporate this critical population size in models of honey bees (Kang et al. 2016; Ratti et al. 2012, 2017; Eberl et al. 2010; Britton and White 2021), leaf-cutter ants (Kang et al. 2011) and eusocial insects in general (Kang and Theraulaz 2016). In our model, it produces a strong Allee effect within each of the subsystems defined by F_{S_c} (7) and F_{S_b} (8). Furthermore, the property extends to the full Filippov system (6) because of the particular shape of the boundary Σ, which is a hyperplane crossing the origin, including all points where $G + g - (W + w + G + g)\theta = 0$. Because each pair $\{E^c_+, E^c_-\}$, $\{E^b_+, E^b_-\}$ and $\{E^p_+, E^p_-\}$ of equilibria shares the same fraction of guards $(G + g)/(W + w + G + g)$ (namely $\gamma_c, \gamma_b, \gamma_p$), each pair lies on the same side of the hyperplane Σ (or on Σ, in the case of pseudoequilibria $\{E^p_+, E^p_-\}$). Moreover, due to the relative position of the equilibria given by $\gamma_b < \gamma_c$ (Theorem 3.5), there are two possibilities described in Sect. 4.1: if the pairs $\{E^c_+, E^c_-\}$ and $\{E^b_+, E^b_-\}$ exist, they either lie on the same side of Σ and one of them is *regular* while the other is *virtual* (not approached by trajectories, see Definition 3.1), or they lie on different sides of Σ and they are both virtual, but there are two pseudoequilibria $\{E^p_+, E^p_-\}$ on Σ. In either case, the global structure typical of strong Allee Effects is maintained for trajectories of the Full System (6): one locally stable extinction equilibrium and, under some existence conditions, a lower, unstable equilibrium and an upper, locally stable equilibrium.

The critical population size in this model is a condition on the non-guarding subpopulation $(W+w)$: guards (G and g) do not contribute to worker production (Hammel et al. 2016), and therefore the colony cannot persist without a minimum number of non-guarding workers regardless of the number of guards. Continued replacement in this case effectively acts as an increased mortality rate or harvesting of the minor non-guarding workers, which are known to make Allee effects more severe and colonies less resilient to worker loss (Dennis and Kemp 2016). In fact, a colony unable to escape the crisis mode not only can sustain less non-guarding workers $((W + w) \mid_{E^c_+} < (W + w) \mid_{E^b_+})$ but also requires a higher minimum population of them to persist $((W + w) \mid_{E^c_-} > (W + w) \mid_{E^b_-})$. Thus, the crisis mode has amplified Allee effects with respect to the non-crisis mode. This further illustrates the trade-off in allocating resources to defense at the expense of colony growth or persistence.

5.2 Heterogeneity and adaptive defense

There are known trade-offs associated with heterogeneity in social-insect colonies. For example, modularity in group interactions causes information loss, but may also improve collective decision in complex environments and reduce the transmission of pathogens (Kao and Couzin 2019; Guo et al. 2020). Stingless-bee colonies maintain a heterogeneous population that varies in morphology, age, and behavioral plasticity (Segers et al. 2015; Grüter et al. 2012; Baudier et al. 2019). Tuning each of these factors enables colonies to navigate the trade-off between the efficiency of producing specialized soldiers and the flexibility required to handle unpredictable changes in the environment. We asked how the parameters regulating morphological specialization, age-based polyethism, and behavioral plasticity impact the colony size and task allocation within the colony. Our results show that a heterogeneous colony composition along each of these three axes is beneficial only in a certain range that depends on the other two. Outside of these ranges, colony function and survival may be compromised.

This is consistent with the idea that certain degrees of diversity can be adaptive for colonies. For example, the mix of cognitive phenotypes among honey-bee foragers allows colonies to balance exploration and exploitation of resources (Cook et al. 2020). Mathematical and computational models show that certain mixtures of these phenotypes maximize resource collection, but the optimal proportion depends on the individual task fidelity of workers (Mosqueiro et al. 2017). Similarly, our model shows that changes in the ratio ρ of fixed morphological types (majors and minors) affect the colony's demographics differently according to the individual task flexibility of minors (replacement), as shown by the interaction between ρ and θ in Figs. 3 and 5.

Colonies can increase their major production through changes in larval feeding, and they do if exposed to frequent threats (Segers et al. 2016). However, the fraction of majors in the colony is usually very small ($\approx$ 1–6%) (Grüter et al. 2012; Hammel et al. 2016). Aside from the additional resources required to produce and maintain majors, our results further illustrate demographic trade-offs associated with major production that may contribute to keeping this proportion low. For instance, with realistic parameters, colony size peaks at low proportions of majors ($\rho \approx 5\%$) and then decreases (but not to extinction) when more majors are produced (see Fig. 6, right panel). The peak is a consequence of using a discontinuous model and might not be evident if using continuous equations, but the inverse relation between colony size and major soldier production (after some proportion) should hold in general because guards do not contribute to brood care. If majors mature into guards at a faster rate, then the colony size peaks at a lower proportion of majors, and high major proportions do cause extinction (Fig. 6, left panel). These effects favor short-term emergency replacement by minors, as opposed to the potential negative long-term demographic consequences of producing more or faster-maturing majors, unless there are chronic, sustained pressures.

Our theoretical results also show the interaction between individual plasticity and morphological specialization in the colony's worker allocation to defense. Figure 5 shows that increasing the production of majors, who later specialize in defense, can actually have a negative impact on the fraction of the colony dedicated to guarding (see Theorem 3.5). This is the case if the colony relies on replacement minor guards to satisfy a minimum required guard proportion (due to either large major death rates or slow maturation of majors), and reduces the available pool of minors by increasing the production of majors. This means that a shift in the ratio of morphological types favoring the production of large soldiers, as observed experimentally for chronically threatened colonies (Segers et al. 2016), is only beneficial if guarding by minors is a transient process and not a long-term requirement for colony defense, which is the case in natural colonies (Baudier et al. 2019).

We also asked how plastic minors should be to allow for a transient recovery from guard loss without maintaining an inefficient body of minor guards. As shown in Fig. 3, for the reference parameter values, if minors were more plastic or the colony demand for guards were to increase (higher θ value), the colony could even go extinct. Thus, minors must be flexible enough to cover the temporary demand for guards in a crisis (see "Non-crisis" trajectory in Fig. 4), but not so sensitive to guard loss that they become a permanent inefficient guarding force ("Crisis" trajectory in Fig. 4, which is not observed in natural colonies) or, in an extreme case, that there are not enough workers taking care of brood and the colony goes extinct (see "Extinction" trajectory in Fig. 4).

5.3 A discontinuous framework

The Filippov framework is certainly not the only alternative for modeling emergency task switches during crisis response in this system or similar ones. Task allocation is modulated by social interactions in social insects (Beshers and Fewell 2001; Beshers et al. 2001; Kang and Theraulaz 2016; Naug and Gadagkar 1999), which can use information from local encounters to implement density-dependent changes in behavior (Gordon et al. 1993; Gordon and Mehdiabadi 1999; Pratt 2005; Gordon et al. 2008; Farji-Brener et al. 2010; Baudier and Pavlic 2020). Thus, it is reasonable to assume that the net replacement rate is a decreasing function of the per-capita probability of minors encountering guards (which could be used to infer the colony demand for defensive tasks), with a threshold below which replacement becomes more likely than not. Response thresholds may vary across individuals and even change with time or experience (Beshers and Fewell 2001; Theraulaz et al. 1998), but we do not model this here except for the difference between majors and minors in performing guarding tasks. Although several continuous functions could approximate the decreasing probability of replacement, we have simplified it to a binary step: all minors are either available for replacement or they are not. Note, however, that the change in behavior modeled by the step function is not replacement itself (which occurs at rate δ, not instantly and simultaneously for all minors) but the availability of minors for replacement. In any case, the assumption that all minors switch their behavior in response to a trigger (guard loss below a threshold) instantly and at the same 'trigger intensity' (θ) represents a limitation of the model.

Although using a discrete replacement function is a strong assumption about the behavior of minors, we believe that a discrete framework can be a valuable lens for systems that have distinct modes of operation. Guard replacement is a crisis response that minor workers only engage in under very specific circumstances. It is different from task switching between routine tasks that are normally included in the worker's repertoire. Thus, we have chosen to model this system's crisis mode and non-crisis mode as two different ways of functioning, with one of them activated only in emergency cases. However, continuous alternatives for the replacement activation could be considered.

A well known continuous function for task switching is given by the Fixed Threshold Model (FTM) (Bonabeau et al. 1996) (reviewed in (Beshers and Fewell 2001)). It proposes a Hill function (analogous to one used for the eclosion term in Eq. (1)) to represent the probability of an individual of cast i switching to a task per unit time as $P_i = s^2/(s^2 + \theta_i^2)$. Here s is the magnitude of the stimulus level corresponding to the task, θ_i is the cast-specific response threshold for the task, and the cast i may correspond to genetic or morphological types, like the "majors" and "minors" in our model. The authors describe the choice of this particular functional form as arbitrary, given that any threshold function should produce similar qualitative results (Bonabeau et al. 1996). We choose the simplest possible threshold function for our model, representing the stimulus level s as the probability of encountering a guard and a minor-specific task threshold of θ. Note that, while the switching probability in the FTM *increases* with s, our switching probability *decreases* with the fraction of guards in the colony. This follows the rationale of social inhibition (Beshers et al. 2001), where encounters signal the *lack* of need for a task, as opposed to models where interactions with a task group increase the switching probability to that task (Kang and Theraulaz 2016).

While using a discontinuous derivative requires additional definitions with respect to continuous dynamical system's analysis, piecewise smooth systems are relatively well understood and used in engineering applications and control theory (Cortes 2008; Di Bernardo et al. 2008; Wang et al. 2019; da Silveira Costa and Meza 2006). A continuous form like the FTM in the derivative would eliminate some mathematical complexities associated with discontinuous derivatives, but would also produce considerably more complex analytical forms for the system's equilibria. Moreover, the exponent in the Hill function proposed by the FTM and similar models would introduce an additional parameter to be estimated.

Recent work has compared the use of Hill functions and Heaviside step functions for a threshold-dependent behavior in social insects (Wang et al. 2022). In that study, one of the key findings is that the continuous model's steady state always has persistence of all the population compartments, unlike the discontinuous model. Although we have not analyzed a continuous version of our model in detail, we have included a preliminary numerical simulation using a Hill function instead of a Heaviside function in Appendix D. Figure 8 illustrates that a continuous replacement function would produce a persistent population of minor guards under non-crisis conditions. This is not what is observed in real colonies, where minors do not perform guarding tasks except for crisis situations (Baudier et al. 2019). A thorough comparison of a discrete and continuous version of this model would constitute informative future work in order to determine the impact of the degree of response nonlinearity on the colony's demographics and task allocation.

5.4 Future work and broader impact

This work provided a theoretical exploration regarding how morphological specialization, age-based polyethism and behavioral plasticity interact at different timescales to regulate group defense and colony growth. To model each of these processes, we have made simplifying assumptions that could be relaxed to produce more biologically realistic models. For instance, we have assumed that there are two distinct morphological types (majors and minors) based on the bimodal distribution of body sizes in stingless bees (Grüter et al. 2012), but this distribution could be explicitly incorporated into a future model to include a range of body sizes. We have also modeled age-based polyethism only for majors, when in reality minors also switch from in-nest to foraging tasks as they mature. If guard replacement is limited to mature minors, which would be consistent with the transition to outside tasks with age (Hammel et al. 2016), then modeling minor aging could elucidate additional interactions between development and behavioral plasticity. Furthermore, as discussed in the previous section, we have not included individual variation in response thresholds for different tasks (Beshers and Fewell 2001) besides the fact that majors always become guards and minors only do so in crisis.

Another set of limitations arises from the availability of data to calibrate the parameter values in the model. In particular, the values of b (the number of brood-caring workers required for half of the eggs laid by the queen to eclose as adults) and θ (the proportion of guards under which minors engage in replacement) were unavailable from the literature (see Table 2). We indirectly computed the former from reports about the number of workers in newly founded colonies (Van Veen and Sommeijer 2000), but no formal estimation was performed. For the latter, we varied θ across its range and analyzed each resulting scenario, but further research into the detection of guard loss by minors would provide a valuable reference for modeling the plasticity of their behavior.

Lastly, we have focused exclusively on task allocation and colony size, but a richer analysis of the balance between reproduction and defense could be obtained by conducting a study of the risks that the colony engages in for being unprotected. This includes quantifying the cost of reducing the guard population in terms of potential attacks or robbing, which are common for stingless bees (Baudier et al. 2019; Segers et al. 2016). Moreover, the indirect contribution of guards to reproduction could be modeled by making the mortality of non-guards inversely proportional to the guard population, as in the aphid soldier-production model by Aoki and Kurosu (2003).

We have presented a mathematical framework for understanding the demographic factors constraining collective defense regulation at different timescales and specialization degrees in social insect colonies. The scenarios shown here shed light on the regulation of specialization and crisis response in colonies, but these results can also be adapted to understand variation at different timescales in the defensive dynamics of other biological systems that are harder to observe. For example, immune systems are also collective distributed systems that must deploy fast responses to crisis situations, and for which some parallels with social-insect colonies have already been drawn (Moses et al. 2019).

Furthermore, efficient resource allocation and effective defense against threats are problems relevant to human societies as well as social-insect colonies. Understanding how social systems with such ecological success address these challenges has the potential to inspire solutions in the human domain. For example, studying the role of heterogeneity of workers within colonies may shed light on the importance of diversity in teams for group problem solving in humans. On the other hand, the dynamics of defensive allocation in stingless bees have already inspired algorithms for robots with guarding tasks (Strickland et al. 2019). In line with this, our work helps to establish social insects as model organisms to understand other systems where the transaction costs for component turnover are nontrivial, as in manufacturing systems and just-in-time supply chains, and thus guide the design of solutions in the human domain.

Acknowledgements This work was partially supported by contract number W31P4Q18-C-0054 from the United States Defense Advanced Research Projects Agency (DARPA). This research of Y.K. is partially supported by National Science Foundation (NSF) - Division of Mathematical Sciences (DMS) (Award Number 1716802&2052820); National Science Foundation-Division of Integrated Organismal Systems (IOS) /Division of Mathematical Sciences (Award Number 1558127) and The James S. McDonnell Foundation 21st Century Science Initiative in Studying Complex Systems Scholar Award (UHC Scholar Award 220020472).

Author Contributions MGNZ, YK, KMB, JHF and TPP contributed to the study conception and design. Analysis was performed by MGNZ and YK. The first draft of the manuscript was written by MGNZ and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials no data were collected in this study.

Declarations

Conflict of interests the authors declare no conflict of interests.

Consent for publication all authors have provided approval for publication.

Code availability No custom algorithms central to the research were used.

Appendix A Summary of system dynamics

Table 3 summarizes the dynamics of Full System (6).

Table 3 Summary of the model's dynamics: equilibrium points for the non-crisis and crisis modes (i.e., without and with guard replacement by minors), including existence conditions (E.C) and local stability (L.S.)

Non-crisis mode without replacement

$$\theta < \frac{G^* + g^*}{P^*} = \frac{\rho\mu_w}{\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau)}$$

	Extinction	Interior	
	E^e	E_+	E_-
W^*	0	$\frac{1}{2}\left(\Lambda\rho\tau + \sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b\tau\mu_w}{\frac{1-\rho}{\rho} + \tau\mu_w}\right)^2}\right)$	$\frac{1}{2}\left(\Lambda\rho\tau - \sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b\tau\mu_w}{\frac{1-\rho}{\rho} + \tau\mu_w}\right)^2}\right)$
G^*	0	$\frac{1}{\tau\mu_G}W^*$	$\frac{1}{\tau\mu_G}W^*$
w^*	0	$\frac{(1-\rho)}{\rho\tau\mu_w}W^*$	$\frac{(1-\rho)}{\rho\tau\mu_w}W^*$
g^*	0	0	0
E.C.	None	$\frac{\Lambda}{2b}(\rho\tau + (1-\rho)\frac{1}{\mu_w}) \geq 1$	
L.S.	Stable	Stable	Unstable

Crisis mode with replacement

$$\theta \geq \frac{G^* + g^*}{P^*} = \frac{\rho\,(\mu_w + \delta) + \mu_G(1-\rho)\frac{\delta}{\mu_g}}{\mu_G(1-\rho)\left(\frac{\delta}{\mu_g} + 1\right) + \rho\,(\mu_w + \delta)\,(1 + \tau\mu_G)}$$

	Extinction	Interior	
	E^e	E_+	E_-
W^*	0	$\frac{1}{2}\left(\Lambda\rho\tau + \sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b\tau(\mu_w+\delta)}{\frac{1-\rho}{\rho} + \tau(\mu_w+\delta)}\right)^2}\right)$	$\frac{1}{2}\left(\Lambda\rho\tau - \sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b\tau(\mu_w+\delta)}{\frac{1-\rho}{\rho} + \tau(\mu_w+\delta)}\right)^2}\right)$

Table 3 continued

Crisis mode with replacement

$$\theta \geq \frac{G^* + g^*}{P^*} = \frac{\rho\,(\mu_w + \delta) + \mu_G(1-\rho)\frac{\delta}{\mu_g}}{\mu_G(1-\rho)\left(\frac{\delta}{\mu_g} + 1\right) + \rho\,(\mu_w + \delta)\,(1 + \tau\mu_G)}$$

	Extinction	Interior	
	E^e	E_+	E_-
G^*	0	$\frac{1}{\tau\mu_G} W^*$	$\frac{1}{\tau\mu_G} W^*$
w^*	0	$\frac{(1-\rho)}{\rho\tau(\mu_w + \delta)} W^*$	$\frac{(1-\rho)}{\rho\tau(\mu_w + \delta)} W^*$
g^*	0	$\frac{\delta(1-\rho)}{\mu_g\rho\tau(\mu_w + \delta)} W^*$	$\frac{\delta(1-\rho)}{2\mu_g\rho\tau(\mu_w + \delta)} W^*$
E.C.	None	$\frac{\Lambda}{2b}(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta}) \geq 1$	
L.S.	Stable	Stable	Unstable

Sliding mode

$\gamma_b < \theta < \gamma_c$

	E_+^p	E_-^p
W^*	$\frac{1}{2}\rho\tau\left(\Lambda + \sqrt{\Lambda^2 - \left(\frac{2b\mu_G(\theta\mu_g + (1-\theta)\mu_w)}{(1-\theta)(\mu_G(1-\rho) + \mu_g\rho + \mu_G\mu_w\rho\tau)}\right)^2}\right)$	$\frac{1}{2}\rho\tau\left(\Lambda - \sqrt{\Lambda^2 - \left(\frac{2b\mu_G(\theta\mu_g + (1-\theta)\mu_w)}{(1-\theta)(\mu_G(1-\rho) + \mu_g\rho + \mu_G\mu_w\rho\tau)}\right)^2}\right)$

Table 3 continued

Sliding mode $\gamma_b < \theta < \gamma_c$		
	E_+^p	E_-^p
G^*	$\frac{1}{\tau \mu_G} W^*$	$\frac{1}{\tau \mu_G} W^*$
w^*	$\frac{(1-\theta)(\mu_G(1-\rho) - \mu_g \rho) + \theta \mu_G \mu_g \rho \tau}{\mu_G \rho \tau (\theta \mu_g + \mu_w (1-\theta))} W^*$	$\frac{(1-\theta)(\mu_G(1-\rho) - \mu_g \rho) + \theta \mu_G \mu_g \rho \tau}{\mu_G \rho \tau (\theta \mu_g + \mu_w (1-\theta))} W^*$
g^*	$\frac{-\mu_w \rho + \theta(\mu_G(1-\rho) + \mu_w \rho (1 + \mu_G \tau))}{\mu_G \rho \tau (\theta \mu_g + \mu_w (1-\theta))} W^*$	$\frac{-\mu_w \rho + \theta(\mu_G(1-\rho) + \mu_w \rho (1 + \mu_G \tau))}{\mu_G \rho \tau (\theta \mu_g + \mu_w (1-\theta))} W^*$
E.C.	$\frac{\Lambda}{2b} \frac{(1-\theta)(\mu_g \rho + \mu_G(1-\rho) + \mu_G \mu_w \tau \rho)}{\mu_G (\theta \mu_g + \mu_w (1-\theta))} \geq 1$	
L.S.	Stable*	Unstable*

*The local stability of the pseudoequilibria $E_\pm^p$ was shown only numerically, not analytically

Appendix B Mathematical Proofs

B.1 Proof of Theorem 3.1 (Basic dynamical properties)

Theorem (Basic dynamical properties) *There exists a unique forward solution for System (6) starting form each initial condition in domain $\mathbb{R}_+^4$. Moreover, $\mathbb{R}_+^4$ is positively invariant with respect to System (6) and System (6) is bounded in $\mathbb{R}_+^4$.*

Proof First, we show the existence and uniqueness of forward solutions for System (6) in $\mathbb{R}_+^4$, following Filippov (1988) (Theorem 2, section 10).

Consider the vector field $F : \mathbb{R}_+^4 \to \mathbb{R}_+^4$ defined by Eq. (9):

$$F(Z) = \begin{cases} F_{S_c}(Z), & Z \in S_c, \\ F_{S_b}(Z), & Z \in S_b, \end{cases} \tag{B1}$$

where $S_c = \left\{Z \in \mathbb{R}_4^+ \mid H(Z) < 0\right\}$, $S_b = \left\{Z \in \mathbb{R}_+^4 \mid H(Z) > 0\right\}$ for

$$H(Z) = (G + g) - \theta(W + G + w + g).$$

F is piecewise continuous with continuous derivatives with respect to W, G, w and g in S_c and S_b up to the boundary, which is defined by smooth surface

$$\Sigma = \left\{Z \in \mathbb{R}_+^4 \mid H(Z) = 0\right\} \in C^\infty.$$

Moreover,

$$F_{S_c} - F_{S_b} = \{0, 0, -\delta w, \delta w\}$$

is continuously differentiable on Σ.

Now we show that, for any point $Z \in \Sigma$, either $F_{S_b}(Z)$ points in the direction of S_c or $F_{S_c}(Z)$ points in the direction of S_b. For $Z \in \Sigma$, define $F_c^N(Z)$ and $F_b^N(Z)$ as the projections of $F_{S_c}(Z)$ and $F_{S_b}(Z)$ onto the normal to Σ in the direction of region S_b, namely

$$H_z := \{-\theta, 1 - \theta, -\theta, 1 - \theta\}.$$

We must show that, for each $Z \in \Sigma$, either $F_c^N(Z) > 0$ or $F_b^N(Z) < 0$. Note that

$$F_c^N - F_b^N = \langle H_z, F_c\rangle - \langle H_z, F_b\rangle = \delta w \geq 0$$

so $F_c^N \geq F_b^N$, and thus we never have $F_c^N \leq 0 \leq F_b^N$ unless $F_c^N = F_b^N = 0$, which is only true for the origin. However, there exists a unique solution starting from the origin: the equilibrium solution. Therefore, by Theorem 2, section 10, of Filippov (1988), there exists a unique forward solution for System (6) starting form each initial condition in domain $\mathbb{R}_+^4$.

Now we show that trajectories with initial conditions in $\mathbb{R}^4_+$ do not become negative or unbounded. Suppose $(W, G, w, g) \in \mathbb{R}^4_+$ is the initial condition for a trajectory. Without loss of generality, assume that W reaches 0 while $w, G, g > 0$. Then

$$\left.\frac{\mathrm{d}W}{\mathrm{d}t}\right|_{W=0} = \Lambda \frac{w^2}{b^2 + w^2}\rho \geq 0$$

so the trajectory cannot escape to a region where $W < 0$. The same holds for $G = 0$ while $W, w, g > 0$, since

$$\left.\frac{\mathrm{d}G}{\mathrm{d}t}\right|_{G=0} = \frac{1}{\tau} W \geq 0$$

and for $w = 0$ while $W, G, g > 0$, because

$$\left.\frac{\mathrm{d}w}{\mathrm{d}t}\right|_{w=0} = \Lambda \frac{W^2}{b^2 + W^2}(1 - \rho) \geq 0.$$

Lastly, for $g = 0$ and $W, w, G > 0$,

$$\left.\frac{\mathrm{d}g}{\mathrm{d}t}\right|_{g=0} = \left[\frac{G}{P} \leq \theta\right] \delta w = \begin{cases} 0\,, & \frac{G}{P} > \theta, \\ \delta w\,, & \frac{G}{P} \leq \theta \end{cases} \geq 0$$

so g cannot become negative.

We can follow a similar procedure for two variables reaching 0 at the time. If $W = G = 0$ and $w, g > 0$, then $\frac{\mathrm{d}W}{\mathrm{d}t} \geq 0$ and $\frac{\mathrm{d}G}{\mathrm{d}t} = 0$, so W and G remain nonnegative. Other combinations can be checked in an analogous manner, as well as cases where three variables reach 0 at the same time. In the case that $W = G = w = g = 0$, then $\frac{\mathrm{d}W}{\mathrm{d}t} = \frac{\mathrm{d}G}{\mathrm{d}t} = \frac{\mathrm{d}w}{\mathrm{d}t} = \frac{\mathrm{d}g}{\mathrm{d}t} = 0$, so no trajectories can escape $\mathbb{R}^4_+$ through the origin. It follows that $\mathbb{R}^4_+$ is positively invariant with respect to System (6).

To show boundedness, we first prove that W is bounded:

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \Lambda \frac{(W + w)^2}{b^2 + (W + w)^2}\rho - \frac{1}{\tau} W \leq \Lambda\rho - \frac{1}{\tau} W.$$

This yields

$$\limsup_{t\to\infty} W \leq \Lambda\rho\tau.$$

Now let $V = W + G + w + g$ and $\mu = \min\{\mu_G, \mu_w, \mu_g\}$. Then we have

$$\begin{aligned} \frac{\mathrm{d}V}{\mathrm{d}t} &= \Lambda \frac{(W + w)^2}{b^2 + (W + w)^2} - (\mu_G G + \mu_w w + \mu_g g) \\ &\leq \Lambda - \mu V + \mu W \leq \Lambda + \mu\Lambda\rho\tau - \mu V, \end{aligned}$$

which gives $\limsup_{t\to\infty} V \leq \Lambda/\mu + \Lambda\rho\tau$. $\square$

B.2 Proof of Theorem 3.2 (extinction equilibrium)

Theorem *Model (6) always has the extinction equilibrium*

$$E^e := (W^*, G^*, w^*, g^*) = (0, 0, 0, 0),$$

which is locally stable. If $\mu_u = \min\{1/\tau, \mu_w\}$ *and* $\Lambda < 2\mu_u b$*, then* E^e *is globally stable.*

Proof When the population is zero, $W = w = G = g = 0$ and

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \frac{\mathrm{d}w}{\mathrm{d}t} = \frac{\mathrm{d}G}{\mathrm{d}t} = \frac{\mathrm{d}g}{\mathrm{d}t} = 0.$$

This means that the system has an extinction equilibrium

$$E^e := (W^*, G^*, w^*, g^*) = (0, 0, 0, 0).$$

We can analyze the stability of this point by studying the linearized system near the origin. It is necessary to evaluate two cases of the Jacobian matrix, one for each case of $[\frac{G+g}{P} \leq \theta]$. The Jacobian matrix of the system can be reduced to

$$\begin{pmatrix} \dfrac{2b^2W^*}{\tau} - \dfrac{1}{\tau} & 0 & \dfrac{2b^2W^*}{\tau} & 0 \\ \dfrac{1}{\tau} & -\mu_G & 0 & 0 \\ \dfrac{2b^2(1-\rho)W^*}{\rho\tau} & 0 & \dfrac{2b^2(1-\rho)W^*}{\rho\tau} - \mu_w & 0 \\ 0 & 0 & 0 & -\mu_g \end{pmatrix}$$

when $\frac{G+g}{P} \leq \theta$, and

$$\begin{pmatrix} \dfrac{2b^2W^*}{\tau} - \dfrac{1}{\tau} & 0 & \dfrac{2b^2W^*}{\tau} & 0 \\ \dfrac{1}{\tau} & -\mu_G & 0 & 0 \\ \dfrac{2b^2(1-\rho)W^*}{\rho\tau} & 0 & \dfrac{2b^2(1-\rho)W^*}{\rho\tau} - \mu_w - \delta & 0 \\ 0 & 0 & \delta & -\mu_g \end{pmatrix}$$

when $\frac{G+g}{P} > \theta$. Evaluating at E^e, we get

$$\begin{pmatrix} -\frac{1}{\tau} & 0 & 0 & 0 \\ \frac{1}{\tau} & -\mu_G & 0 & 0 \\ 0 & 0 & -\mu_w & 0 \\ 0 & 0 & 0 & -\mu_g \end{pmatrix} \text{ and } \begin{pmatrix} -\frac{1}{\tau} & 0 & 0 & 0 \\ \frac{1}{\tau} & -\mu_G & 0 & 0 \\ 0 & 0 & -\mu_w - \delta & 0 \\ 0 & 0 & \delta & -\mu_g \end{pmatrix}$$

respectively. The eigenvalues of these matrices are

$$\left\{-\frac{1}{\tau}, -\mu_g, -\mu_G, -\mu_w\right\} \text{ and } \left\{-\frac{1}{\tau}, -\mu_g, -\mu_G, -\mu_w - \delta\right\}$$

which are all negative values. This means that the zero equilibrium is locally stable, regardless of the value of $\frac{G+g}{P}$ in its vicinity.

Now assume that $\Lambda < 2\mu_u b$ for $\mu_u = \min\{1/\tau, \mu_w\}$. We will show that E^e is globally asymptotically stable.

Let

$$u = W + w.$$

Then

$$\begin{aligned} \frac{du}{dt} &= \Lambda \frac{(W+w)^2}{b^2 + (W+w)^2} - W/\tau - \mu_w w \\ &< \Lambda \frac{u^2}{b^2+u^2} - \mu_u u = u \frac{\Lambda u - \mu_u (b^2 + u^2)}{b^2 + u^2} \\ &= \mu_u u \frac{\phi}{b^2+u^2} \end{aligned}$$

for

$$\phi = -\left(u - \frac{\Lambda}{2\mu_u}\right)^2 + \frac{\left(\frac{\Lambda}{\mu_u}\right)^2 - 4b^2}{4}.$$

Because $\Lambda/\mu_u < 2b$, ϕ is negative. Therefore, both W and w collapse to zero if $\Lambda/\mu_u < 2b$. In this case,

$$\frac{\mathrm{d}G}{\mathrm{d}t} = -\mu_G G \text{ and } \frac{\mathrm{d}g}{\mathrm{d}t} = -\mu_g g.$$

Therefore, G and g also collapse to extinction. This completes the proof. □

B.3 Proof of Theorem 3.3 (existence and stability of the crisis interior equilibria)

Theorem *Define condition*

$$C_c : \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w+\delta}\right) \geq 1.$$

Then the Crisis System defined by F_{S_c} (7) *has two interior equilibria,* E^c_+ *and* E^c_-*, if and only if* C_c *holds. Both have the form*

$$(W^*, G^*, w^*, g^*) = (W^*, \frac{1}{\tau\mu_G}W^*, \frac{1-\rho}{\rho\tau(\mu_w+\delta)}W^*, \frac{\delta}{\mu_g}\frac{1-\rho}{\rho\tau(\mu_w+\delta)}W^*)$$

where

$$W^* = \frac{1}{2}\left(\Lambda\rho\tau \pm \sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b(\mu_w+\delta)\tau}{\frac{1-\rho}{\rho} + (\mu_w+\delta)\tau}\right)^2}\right) \tag{B2}$$

Moreover,

1. *The interior equilibrium* E^c_+ *is Locally Asymptotically Stable, and* E^c_- *is unstable.*
2. *Both* E^c_+ *and* E^c_- *are* regular *equilibria of the Full System* (6) *if and only if*

$$\gamma_c < \theta$$

for γ_c *defined in Eq.* (11).

Proof From relations (10), each equilibrium of System (7) takes the form:

$$(W^*, G^*, w^*, g^*) = (W^*, \frac{1}{\tau\mu_G}W^*, \frac{1-\rho}{\rho\tau(\mu_w+\delta)}W^*, \frac{\delta}{\mu_g}\frac{1-\rho}{\rho\tau(\mu_w+\delta)}W^*).$$

Furthermore, the value of W^* in an interior equilibrium of the System (7) must be a root of Eq. (12), which can be rearranged as:

$$f(W^*) = W^{*2} - \Lambda\tau\rho W^* + (\frac{b(\mu_w+\delta)\tau}{\frac{1-\rho}{\rho} + (\mu_w+\delta)\tau})^2.$$

with all parameters being positive and finite, and that root must be positive by Theorem 3.1. It follows that there are at most two interior equilibria, E^c_+ and E^c_-, and those equilibria are such that:

$$W^* = \frac{1}{2}\left(\Lambda\rho\tau \pm \sqrt{\underbrace{(\Lambda\rho\tau)^2 - \left(\frac{2b(\mu_w+\delta)\tau}{\frac{1-\rho}{\rho} + (\mu_w+\delta)\tau}\right)^2}_{*}}\right),$$

Thus condition C_c of this theorem, equivalent to the discriminant ($*$ above) being positive, holds if and only if the two positive, real roots described by (B4) exist.

1. We study the local stability of the equilibrium points using the Jacobian matrix J of the linearized system in their vicinity. From System (7), we get

$$J = \begin{pmatrix} \frac{2b^2\Lambda\rho(w+W)}{\left(b^2+(w+W)^2\right)^2} - \frac{1}{\tau} & 0 & \frac{2b^2\Lambda\rho(w+W)}{\left(b^2+(w+W^*)^2\right)^2} & 0 \\ \frac{1}{\tau} & -\mu_G & 0 & 0 \\ -\frac{2b^2\Lambda(\rho-1)(w^*+W^*)}{\left(b^2+(w^*+W^*)^2\right)^2} & 0 & -\frac{2b^2\Lambda(\rho-1)(w\ +W\ )}{\left(b^2+(w\ +W\ )^2\right)^2} - \mu_w - \delta & 0 \\ 0 & 0 & \delta & -\mu_g \end{pmatrix}$$

From the characteristic polynomial, and using Eqs. (7) to obtain an equation in terms of W, it follows that the eigenvalues of J satisfy the equation

$$(\mu_g + \lambda)(\mu_G + \lambda)\left(\lambda^2 + \lambda B + C\right)$$

where

$$\begin{aligned} B &:= \frac{1}{\tau} + \mu_w + \delta - h, \\ &= \frac{W}{(W+w)\left(b^2+(W+w)^2\right)\tau\rho}\left(\rho\frac{w}{W}\left(b^2+(W+w)^2\right)\right. \\ &\quad \left. + (1-\rho)\frac{W}{w}\left(b^2+(W+w)^2\right) + (W+w+b)(W+w-b)\right), \\ C &:= \frac{1}{\tau}(\mu_w+\delta) - \rho h(\mu_w+\delta) - \frac{1}{\tau}(1-\rho)h, \\ &= \frac{\frac{(\mu_w+\delta)}{\tau}W}{b^2+(W+w)^2}f'(W), \end{aligned}$$

where $h := (2b^2\Lambda(W+w))/(b^2+(W+w)^2)^2$. This yields the two negative eigenvalues $-\mu_g$ and $-\mu_G$, and two additional ones, λ_+ and λ_-, that satisfy a quadratic equation. Define the sum and product of λ_+ and λ_- as

$$\Pi := \lambda_+\lambda_- = C$$

$$\Sigma := \lambda_+ + \lambda_- = -B.$$

With this terminology, we analyze the sign of λ_+ and λ_- for each of the interior equilibria E_+^c and E_-^c.

From the existence condition of equilibrium we have that

$$W + w - b = \frac{1}{2}\left(1 + \frac{1-\rho}{\tau\rho(\mu_w + \delta)}\right)\sqrt{(\Lambda\rho\tau)^2 - \left(\frac{2b(\mu_w + \delta)\tau}{\frac{1-\rho}{\rho} + (\mu_w + \delta)\tau}\right)^2}$$
$$+ \frac{1}{2}\left(1 + \frac{1-\rho}{\tau\rho(\mu_w + \delta)}\right)\tau\rho\Lambda - b > 0.$$

Thus, we have $B > 0$. Because $g'(W) < 0$ at equilibrium E_-^c and $g'(W) > 0$ at equilibrium E_+^c, then we have $C|_{E_-^c} < 0$ and $C|_{E_+^c} > 0$. According to the Routh–Hurwitz condition, we can get that E_-^c is always unstable and E_+^c is locally asymptotically stable.

2. By Definition 3.1, Z^* is a regular equilibrium of System (6) iff $F_{S_c}(Z^*) = 0$, $H(Z^*) < 0$ or $F_{S_b}(Z^*) = 0$, $H(Z^*) > 0$, with F_{S_c} and F_{S_b} defined in Eqs. (7) and (8), and $H(Z) = (G + g) - \theta P$. By construction,

$$F_{S_c}(E_+^c) = F_{S_c}(E_-^c) = 0.$$

Moreover, note that both E_-^c and E_+^c have the same guard ratio $(G^* + g^*)/P^*$, defined as γ_c in Eq. (11). It follows that

$$H(E_-^c) < 0 \iff H(E_+^c) < 0 \iff \gamma_c < 0.$$

Thus, both E_-^c and E_+^c are regular equilibria of System (6) iff $\gamma_c < \theta$.

□

B.4 Proof of Theorem 3.4 (existence and stability of the non-crisis interior equilibria)

Theorem *Define condition*

$$C_b : \frac{\Lambda}{2b}\left(\rho\tau + (1-\rho)\frac{1}{\mu_w}\right) \geq 1. \tag{B3}$$

Then the Non-Crisis System defined by F_{S_b} (8) has two interior equilibria $E_+^b = E_+^c|_{\delta=0}$ and $E_-^b = E_-^c|_{\delta=0}$ if and only if C_b holds.

Moreover,

1. *The interior equilibrium E_+^b is Locally Asymptotically Stable, while E_-^b is unstable.*

2. *Both* E_+^b *and* E_-^b *are* regular *equilibria of the Full System* (6) *if and only if*

$$\gamma_b > \theta$$

for γ_b *defined in Eq.* (15).

Proof The Non-Crisis System F_{S_b} (8) is identical to the Crisis System F_{S_c} (7) when $\delta = 0$. Thus, following the same analysis as the previous proof in Sect. B.3 in the special case $\delta = 0$, we get that F_{S_b} (8) has two positive, real equilibria E_+^b and E_-^b, of the form

$$(W^*, G^*, w^*, g^*) = (W^*, \frac{1}{\tau\mu_G} W^*, \frac{1-\rho}{\rho\tau\mu_w} W^*, 0)$$

with

$$W^* = \frac{1}{2}\left(\Lambda\rho\tau \pm \sqrt{\underbrace{(\Lambda\rho\tau)^2 - \left(\frac{2b\mu_w\tau}{\frac{1-\rho}{\rho} + \mu_w\tau}\right)^2}_{*}}\right). \tag{B4}$$

if and only if the discriminant ($*$ above) is guaranteed to be positive or, equivalently, condition C_b of this theorem holds.

1. We study the local stability of the equilibrium points using the Jacobian matrix J of the linearized system in their vicinity. From System (8), we get

$$J = \begin{pmatrix} \frac{2b^2\Lambda\rho(w+W)}{\left(b^2+(w+W)^2\right)^2} - \frac{1}{\tau} & 0 & \frac{2b^2\Lambda\rho(w+W)}{\left(b^2+(w+W^*)^2\right)^2} & 0 \\ \frac{1}{\tau} & -\mu_G & 0 & 0 \\ -\frac{2b^2\Lambda(\rho-1)(w^*+W^*)}{\left(b^2+(w^*+W^*)^2\right)^2} & 0 & -\frac{2b^2\Lambda(\rho-1)(w\ +W\ )}{\left(b^2+(w\ +W\ )^2\right)^2} - \mu_w & 0 \\ 0 & 0 & 0 & -\mu_g \end{pmatrix}$$

From the characteristic polynomial, it follows that the eigenvalues of J are roots of

$$(\mu_g + \lambda)(\mu_G + \lambda)\left(\lambda^2 + \lambda B + C\right) = 0$$

where

$$B := \frac{1}{\tau} + \mu_w - h,$$

$$
\begin{aligned}
&= \frac{W}{(W+w)\left(b^2+(W+w)^2\right)\tau\rho}\left(\rho\frac{w}{W}\left(b^2+(W+w)^2\right)\right.\\
&\qquad\left.+(1-\rho)\frac{W}{w}\left(b^2+(W+w)^2\right)+(W+w+b)(W+w-b)\right),\\
C &:= \frac{\mu_w}{\tau}-\mu_w\rho h-\frac{1}{\tau}(1-\rho)h,\\
&= \frac{\frac{\mu_w}{\tau}W}{b^2+(W+w)^2}f'(W),
\end{aligned}
$$

and $h := (2b^2\Lambda(W+w))/(b^2+(W+w)^2)^2$. Consequently, there are two negative eigenvalues $-\mu_g$ and $-\mu_G$, and two additional ones, λ_+ and λ_-, that satisfy a quadratic equation. Define the sum and product of λ_+ and λ_- as

$$
\begin{aligned}
\Pi &:= \lambda_+\lambda_- = C\\
\Sigma &:= \lambda_+ + \lambda_- = -B.
\end{aligned}
$$

With this terminology, we analyze the sign of λ_+ and λ_- for each of the interior equilibria E_+^b and E_-^b.

From the existence condition C_b, we have that

$$
\begin{aligned}
W+w-b &= \frac{1}{2}\left(1+\frac{1-\rho}{\tau\rho\mu_w}\right)\sqrt{(\Lambda\rho\tau)^2-\left(\frac{2b\mu_w\tau}{\frac{1-\rho}{\rho}+\mu_w\tau}\right)^2}\\
&\quad+\frac{1}{2}\left(1+\frac{1-\rho}{\tau\rho\mu_w}\right)\tau\rho\Lambda-b>0.
\end{aligned}
$$

Thus, we have $B > 0$. Because $f'(W) < 0$ at equilibrium E_-^b and $f'(W) > 0$ at equilibrium E_+^b, then we have $C|_{E_-^b} < 0$ and $C|_{E_+^b} > 0$. According to Routh–Hurwitz condition, we get that E_-^b is always unstable and E_+^b is locally asymptotically stable. This completes this proof.

2. By Definition 3.1, Z^* is a regular equilibrium of System (6) iff $F_{S_c}(Z^*) = 0$, $H(Z^*) < 0$ or $F_{S_b}(Z^*) = 0$, $H(Z^*) > 0$, with F_{S_c} and F_{S_b} defined in Eqs. (7) and (8), and $H(Z) = (G+g) - \theta P$. By construction,

$$
F_{S_b}(E_+^b) = F_{S_b}(E_-^b) = 0.
$$

Moreover, note that both E_-^b and E_+^b have the same guard ratio $\dfrac{G^*+g^*}{P^*}$, defined as γ_b in Eq. (15). It follows that

$$
H(E_-^b) < 0 \iff H(E_+^b) < 0 \iff \gamma_b < 0.
$$

Thus, both E_-^b and E_+^b are regular equilibria of System (6) iff $\gamma_b > \theta$. □

B.5 Proof of Theorem 3.5 (characterization of the equilibrium guard ratios)

Theorem *Consider the ratios*

$$\gamma_b = \frac{\rho\mu_w}{\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau)}$$

defined in Eq. (15) *and*

$$\gamma_c = \frac{\rho\,(\mu_w+\delta) + \mu_G(1-\rho)\frac{\delta}{\mu_g}}{\mu_G(1-\rho)\left(\frac{\delta}{\mu_g}+1\right) + \rho\,(\mu_w+\delta)\,(1+\tau\mu_G)}.$$

defined in Eq. (11)*. Then the following are true:*

1. *The ratio γ_b is monotonically increasing with respect to ρ.*
2. *The ratio γ_c is monotonically decreasing with respect to ρ if $\mu_g \leq \delta\tau\mu_G$ while it is monotonically increasing otherwise.*
3. *Moreover, we always have*

$$\gamma_b \leq \gamma_c,$$

which implies that two positive, stable, regular equilibria cannot coexist in the Full System (6).

Proof 1. Consider the derivative of γ_b with respect to ρ:

$$\frac{\partial\gamma_b}{\partial\rho} = \frac{\mu_G\mu_w}{(\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau))^2} > 0$$

Thus, γ_b is a monotonically increasing function of ρ.

2. Similarly, consider the derivative of γ_c with respect to ρ:

$$\frac{\partial\gamma_c}{\partial\rho} = \frac{\mu_g\mu_G\,(\delta+\mu_w)\left(\mu_g - \delta\tau\mu_G\right)}{\left(\mu_G(1-\rho)\left(\frac{\delta}{\mu_g}+1\right) + \rho\,(\mu_w+\delta)\,(1+\tau\mu_G)\right)^2}$$

It follows that

$$\begin{cases} \dfrac{\partial\gamma_c}{\partial\rho} \leq 0\,, & \mu_g \leq \delta\tau\mu_G \\ \dfrac{\partial\gamma_c}{\partial\rho} > 0\,, & \text{otherwise} \end{cases}.$$

3. Moreover,

$$\frac{\rho\mu_w}{\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau)} \leq \frac{\rho(\mu_w+\delta) + \mu_G(1-\rho)\delta/\mu_g}{\mu_G(1-\rho)(1+\delta/\mu_g) + \rho(\mu_w+\delta)(1+\tau\mu_G)}$$

$$\iff \rho\mu_w\mu_G(1-\rho)(1+\delta/\mu_g) + \mu_w\rho^2(\mu_w+\delta)(1+\mu_G\tau)$$
$$\leq (\rho(\mu_w+\delta) + \mu_G(1-\rho)\delta/\mu_g)(\mu_G(1-\rho) + \rho\mu_w(1+\mu_G\tau))$$

After some manipulation, the last inequality can be reduced to

$$-\rho\mu_w\mu_G\tau\delta/\mu_g - \rho\delta - \mu_G(1-\rho)\delta/\mu_g \leq 0$$

which is always true. Thus, $\gamma_b \leq \gamma_c$. By the conditions on θ from Theorems 3.3–3.4, two regular stable equilibria E_+^c and E_+^b can coexist only if $\gamma_c < \theta < \gamma_b$, which we just proved impossible. It follows that the Full System (6) can have at most one positive, stable, regular equilibrium.

This completes the proof. □

B.6 Proof of Theorem 3.6 (existence of pseudoequilibria)

Theorem *Define condition*

$$C_p : \frac{\Lambda}{2b}\frac{(1-\theta)(\mu_g\rho + \mu_G(1-\rho) + \mu_G\mu_w\tau\rho)}{\mu_G(\theta\mu_g + \mu_w(1-\theta))} \geq 1.$$

Then System (6) has two pseudoequilibria E_+^p and E_-^p on the switching surface Σ if and only if C_p holds and $\gamma_b \leq \theta < \gamma_c$. The pseudoequilibria have the form

$$(W^*, G^*, g^*, w^*) = \left(W^*, \frac{1}{\tau\mu_G}W^*, \frac{(1-\theta)(\mu_G(1-\rho) - \mu_g\rho) + \theta\mu_G\mu_g\rho\tau)}{\mu_G\rho\tau(\theta\mu_g + \mu_w(1-\theta))}W^*, \right.$$
$$\left. \frac{-\mu_w\rho + \theta(\mu_G(1-\rho) + \mu_w\rho(1+\mu_G\tau))}{\mu_G\rho\tau(\theta\mu_g + \mu_w(1-\theta))}W^*\right)$$

with

$$W^*|_{E_\pm^p} = \frac{1}{2}\rho\tau\left(\Lambda \pm \sqrt{\Lambda^2 - \left(\frac{2b\mu_G(\theta\mu_g + (1-\theta)\mu_w)}{(1-\theta)(\mu_G(1-\rho) + \mu_g\rho + \mu_G\mu_w\rho\tau)}\right)^2}\right)$$

Proof From relations (21), any pseudoequilibrium of System (6) must take the form:

$$(W^*, G^*, g^*, w^*) = (W^*, \frac{1}{\tau\mu_G}W^*, \frac{1-\rho}{\rho\tau(\mu_w + \delta(1-\lambda^*))}W^*, \frac{\delta(1-\lambda^*)(1-\rho)}{\mu_g\rho\tau(\mu_w + \delta(1-\lambda^*))}W^*)$$

Recall the value of λ^* in Eq. (23)

$$\lambda^* = \frac{\mu_G\mu_g\theta(\delta\tau\rho + 1 - \rho + \mu_w\rho\tau) - (1-\theta)(\delta(\mu_G(1-\rho) + \mu_g\rho) + \mu_g\mu_w\rho)}{\delta(\mu_G\mu_g\rho\tau\theta - (1-\theta)(\mu_G(1-\rho) + \mu_g\rho))} \tag{B5}$$

First, note that $\lambda^* \in [0, 1] \iff \gamma_b \le \theta \le \gamma_c$. In fact, the equation for λ^* can be rearranged as a function of θ as

$$\lambda^*(\theta) = \frac{-A + \theta B}{\delta\,(C + \theta D)} \tag{B6}$$

with

$$\begin{aligned}
A &= \delta(\mu_G(1-\rho) + \mu_g\rho) + \mu_g\mu_w\rho \\
B &= \delta(\mu_G(1-\rho) + \mu_g\rho(1+\mu_G\tau)) + \mu_g(\mu_G(1-\rho) + \mu_w\rho(1+\mu_G\tau)) \\
C &= -\mu_G(1-\rho) + \mu_g\rho \\
D &= \mu_G(1-\rho) + \mu_g\rho(1+\mu_G\tau).
\end{aligned}$$

This is a monotonically decreasing hyperbola, as the derivative

$$\frac{\partial\lambda^*}{\partial\theta} = \frac{-\mu_g\mu_G(1-\rho)\left(\mu_G(1-\rho) + \mu_g\rho + \mu_G\mu_w\rho\tau\right)}{\delta^2\left(-(1-\theta)\mu_g\rho + \mu_G(-(1-\rho) + \theta(1-\rho+\mu_g\rho\tau))\right)^2}$$

is negative for all θ except at the vertical asymptote, where it is undefined. Furthermore, the hyperbola has a horizontal asymptote at

$$1 + \frac{\mu_g\,(\mu_G(1-\rho) + \mu_w\rho(1+\mu_G\tau))}{\delta\left(\mu_G(1-\rho) + \mu_g\rho(1+\mu_G\tau)\right)} > 1.$$

Moreover, $\theta = \gamma_b \implies \lambda^* = 1$ and $\theta = \gamma_c \implies \lambda^* = 0$. Thus,

$$\lambda^* \in [0, 1] \iff \gamma_b \le \theta \le \gamma_c$$

By substituting the value of λ^* from Eq. (23), we get the form

$$\begin{aligned}
(W^*, G^*, g^*, w^*) = \Bigg(W^*, \frac{1}{\tau\mu_G}W^*, \frac{(1-\theta)(\mu_G(1-\rho) - \mu_g\rho) + \theta\mu_G\mu_g\rho\tau)}{\mu_G\rho\tau(\theta\mu_g + \mu_w(1-\theta))}W^*, \\
\frac{-\mu_w\rho + \theta(\mu_G(1-\rho) + \mu_w\rho(1+\mu_G\tau))}{\mu_G\rho\tau(\theta\mu_g + \mu_w(1-\theta))}W^*\Bigg)
\end{aligned}$$

Furthermore, the value of W^* must be a root of

$$f(W^*) = W^{*2} - \Lambda\tau\rho W^* + \Big(\frac{b(\mu_w + \delta(1-\lambda^*))\tau}{\frac{1-\rho}{\rho} + (\mu_w + \delta(1-\lambda^*))\tau}\Big)^2$$

with all parameters being positive and finite, and that root must be positive by Theorem 3.1. This yields the existence condition

$$C_p : \frac{\Lambda}{2b}(\rho\tau + (1-\rho)\frac{1}{\mu_w + \delta(1-\lambda^*)}) \geq 1$$

Substituting λ^* again from (23), it follows that there are at most two pseudoequilibria, E_+^p and E_-^p, such that

$$W^*|_{E_\pm^p} = \frac{1}{2}\rho\tau\left(\Lambda \pm \sqrt{\Lambda^2 - \left(\frac{2b\mu_G(\theta\mu_g + (1-\theta)\mu_w)}{(1-\theta)(\mu_G(1-\rho) + \mu_g\rho + \mu_G\mu_w\rho\tau)}\right)^2}\right)$$

Now assume that $\gamma_b \leq \theta < \gamma_c$. We will show that E_+^p and E_-^p lay on the sliding set $\Sigma_s \subseteq \Sigma$. That is, we will demonstrate $\sigma(E_\pm^p) \leq 0$, for the function σ defined in (18). In fact,

$$\sigma(E_\pm^p) = \phi\frac{\left(\Lambda \pm \sqrt{\Lambda^2 - \left(\frac{2b\mu_G(\theta(\mu_g - \mu_w) + \mu_w)}{(1-\theta)(\mu_G + \mu_g\rho + \mu_G\rho(-1+\mu_w\tau)}\right)^2}\right)^2}{4\mu_G^2(\theta(\mu_g - \mu_w) + \mu_w)^2}$$

with

$$\begin{aligned}
\phi := \theta^2\Big(&\delta\rho\mu_g\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right)\\
&+ \delta\rho\mu_G\left(\tau\mu_g - 1\right)\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right)\\
&+ \mu_g\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right)^2 + \delta\mu_G\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right)\Big)\\
&+ \theta\left(-\delta\rho^2\mu_G\mu_w\left(\tau\mu_g - 1\right) - \delta\rho\mu_g\Big(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\Big)\right.\\
&- 2\rho\mu_g\mu_w\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right) - \delta\rho^2\mu_g\mu_w\\
&+ \delta\rho\mu_G\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right)\\
&\left.- \delta\mu_G\left(\mu_G + \rho\mu_G\left(\tau\mu_w - 1\right) + \rho\mu_w\right) - \delta\rho\mu_G\mu_w\right)\\
&+ \delta\rho^2\mu_g\mu_w + \rho^2\mu_g\mu_w^2 - \delta\rho^2\mu_G\mu_w + \delta\rho\mu_G\mu_w
\end{aligned}$$

It follows that $\sigma(E_\pm^p) \leq 0 \iff \phi \leq 0$. Therefore, we examine the sign of ϕ. As a function of θ, $\phi(\theta)$ is a parabola with two positive roots γ_b and γ_c. Moreover,

$$\phi(0) = \delta\rho^2\mu_g\mu_w + \rho^2\mu_g\mu_w^2 + \delta\rho\mu_G\mu_w(1-\rho) \geq 0$$

so the parabola has a positive intercept and two positive roots. Then it must be negative between its roots, that is, $\phi(\theta) \leq 0$ for $\gamma_b \leq \theta \leq \gamma_c$. Therefore, E_-^p and E_+^p belong to the sliding set Σ_s.

We have shown that E_-^p and E_+^p are indeed pseudoequilibria when $\gamma_b \leq \theta \leq \gamma_c$ and condition C_p holds. We can also see that, should any of these conditions fail to hold, there are no pseudoequilibria. In fact, if $\theta < \gamma_b$ or $\theta > \gamma_c$, then $\lambda^* \notin [0, 1]$ and E_-^p and E_+^p do not lay on the sliding segment Σ_s, and if C_p is false then E_-^p and E_+^p are not real values. This completes the proof. □

Appendix C Parameter values for *T. angustula* stingless bees

General assumptions for stingless bee colonies:

- $\delta \gg 1/\tau$: replacement by minors is a fast response mechanism and occurs at a faster rate than the natural maturation of majors into guards
- $\tau < 1/\mu_w$: maturation of majors is shorter than the average lifespan of minors
- $\mu_g > \mu_G$: minors are worst at defense, so they die sooner thank major guards
- $\rho \ll 1$: majors are the minority of the colony

Specific values found in literature, summarized in Table 2:

- The average guarding duration was estimated to be 5.4 ± 1.5 days (Hammel et al. 2016; Grüter et al. 2011)
- The average age of guarding workers was 26.3 ± 3.3 days, and bees started guarding at 20 days of age (Hammel et al. 2016)
- Replacement guards are observed 5 h after removing guards (Baudier et al. 2019). That is, replacement occurs within 5 h (δ is at least 4/day)
- The final age (last day a bee was seen in the hive) between the two size classes (majors: 27.85 ± 5.6 days; minors: 27.0 ± 8.4 days) does not differ (Hammel et al. 2016). In another study, the average lifespan of bees was estimated to be 21 days (Grüter et al. 2011)
- In a swarming colony estimated to have around 10,000 adults, the recently founded offspring colonies had between 500 and 1000 workers (Van Veen and Sommeijer 2000)
- The egg-laying rate by the queen was on average 6.41 eggs per hour in an observation period of 18 h per day (Koedam et al. 1997)

Appendix D Example comparison to continuous model

Although the analysis of alternative continuous functions for the replacement of guards by minors was not performed in this work, we have included preliminary simulations comparing the use of the step function in Eq. (5) with a Hill function as suggested by the Fixed Threshold Model (Bonabeau et al. 1996) and derivatives thereof. Namely, instead of

$$\left[\frac{G+g}{P} \leq \theta\right] \tag{D1}$$

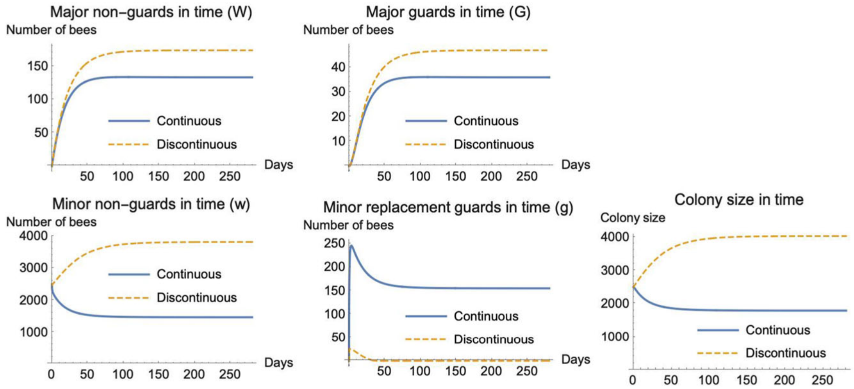

Fig. 8 Simulations comparing the use of a discrete (step) function and a continuous (Hill) function for response of minors to guard loss

we use

$$1 - \frac{(\frac{G+g}{P})^2}{(\frac{G+g}{P})^2 + \theta^2} \tag{D2}$$

for the behavior of minors. In Fig. 8 reproduce the ”Non-crisis” simulation in Fig. 4 comparing the two models: The steady state in the continuous model presents a persistent population of minor guards for realistic parameters, whereas experimental observations of *T. angustula* stingless bees report no minors in guarding tasks unless major guards are removed (Baudier et al. 2019).

References

Aoki S, Kurosu U (2003) Logistic model for soldier production in aphids. Insectes Soc 50(3):256–261. https://doi.org/10.1007/s00040-003-0675-3

Baudier KM, Pavlic TP (2020) Incidental interactions among Neotropical army-ant colonies are met with self-organized walls of ants (Hymenoptera: Formicidae). Myrmecol News 30:251–258. https://doi.org/10.25849/MYRMECOL.NEWS_030:251

Baudier KM, Ostwald MM, Grüter C et al (2019) Changing of the guard: mixed specialization and flexibility in nest defense (Tetragonisca angustula). Behav Ecol 30(4):1041–1049. https://doi.org/10.1093/beheco/arz047

Bellman R, Glicksberg I, Gross O (1956) On the “bang-bang” control problem. Q Appl Math 14(1):11–18

Beshers SN, Fewell JH (2001) Models of division of labor in social insects. Annu Rev Entomol 46(1):413–440. https://doi.org/10.1146/annurev.ento.46.1.413

Beshers SN, Huang ZY, Oono Y et al (2001) Social inhibition and the regulation of temporal polyethism in honey bees. J Theor Biol 213(3):461–479. https://doi.org/10.1006/jtbi.2001.2427

Bonabeau E, Theraulaz G, Deneubourg JL (1996) Quantitative study of the fixed threshold model for the regulation of division of labour in insect societies. Proc R Soc Lond B 263(1376):1565–1569

Boukal DS, Kivan V (1999) Lyapunov functions for Lotka-Volterra predator-prey models with optimal foraging behavior. J Math Biol 39(6):493–517

Britton NF, White JK (2021) The effect of covert and overt infections on disease dynamics in honey-bee colonies. Bull Math Biol 83(6):67

Caetano-Anollés K, Ewers B, Iyer S et al (2021) A minimal framework for describing living systems: a multi-dimensional view of life across scales. Integr Comp Biol 61(6):2053–2065. https://doi.org/10.1093/icb/icab172

Camazine S, Deneubourg JL, Franks NR et al (2003) Self-organization in biological systems. Princeton University Press, Princeton

Cook CN, Lemanski NJ, Mosqueiro T, et al (2020) Individual learning phenotypes drive collective behavior. In: Proceedings of the national academy of sciences of the United States of America 117(30):17,949–17,956. https://doi.org/10.1073/pnas.1920554117, https://figshare.com/articles/dataset/Individual_Learning_Phenotypes_Drive_Collective_Behavior/9775955/6, publisher: Figshare

Cortes J (2008) Discontinuous dynamical systems. IEEE Control Syst Mag 28(3):36–73. https://doi.org/10.1109/MCS.2008.919306

Dennis B, Kemp WP (2016) How hives collapse: Allee effects, ecological resilience, and the honey bee. PLoS ONE 11(2):e0150,055

da Silveira Costa MI, Meza MEM (2006) Application of a threshold policy in the management of multi-species fisheries and predator culling. Math Med Biol J IMA 23(1):63–75

Di Bernardo M, Budd CJ, Champneys AR et al (2008) Bifurcations in nonsmooth dynamical systems. SIAM Rev 50(4):629–701

Dornhaus A (2008) Specialization does not predict individual efficiency in an ant. PLoS Biol. https://doi.org/10.1371/journal.pbio.0060285

Eberl HJ, Frederick MR, Kevan PG (2010) Importance of brood maintenance terms in simple models of the honeybee-varroa destructor-acute bee paralysis virus complex. Electron J Differ Equ (EJDE) [electronic only] 2010:85–98

Farji-Brener AG, Amador-Vargas S, Chinchilla F et al (2010) Information transfer in head-on encounters between leaf-cutting ant workers: food, trail condition or orientation cues? Anim Behav 79(2):343–349. https://doi.org/10.1016/j.anbehav.2009.11.009

Fewell JH (2003) Social insect networks. Science. https://doi.org/10.1126/science.1088945

Filippov AF (1988) Equations with the right-hand side continuous in x and discontinuous in t. In: Differential equations with discontinuous righthand sides. Springer, pp 3–47

Gordon DM (2002) The organization of work in social insect colonies. Complexity. https://doi.org/10.1002/cplx.10048

Gordon DM, Mehdiabadi NJ (1999) Encounter rate and task allocation in harvester ants. Behav Ecol Sociobiol 45(5):370–377. https://doi.org/10.1007/s002650050573

Gordon DM, Paul RE, Thorpe K (1993) What is the function of encounter patterns in ant colonies? Anim Behav 45(6):1083–1100. https://doi.org/10.1006/anbe.1993.1134

Gordon DM, Holmes S, Nacu S (2008) The short-term regulation of foraging in harvester ants. Behav Ecol 19(1):217–222. https://doi.org/10.1093/beheco/arm125

Goutelle S, Maurin M, Rougier F et al (2008) The Hill equation: a review of its capabilities in pharmacological modelling. Funda Clin Pharmacol 22(6):633–648

Grüter C, Kärcher M, Ratnieks F (2011) The natural history of nest defence in a stingless bee, Tetragonisca angustula (Latreille)(Hymenoptera: Apidae), with two distinct types of entrance guards. Neotrop Entomol 40:55–61

Grüter C, Menezes C, Imperatriz-Fonseca VL et al (2012) A morphologically specialized soldier caste improves colony defense in a neotropical eusocial bee. Proc Natl Acad Sci USA 109(4):1182–1186. https://doi.org/10.1073/pnas.1113398109

Guo X, Chen J, Azizi A et al (2020) Dynamics of social interactions, in the flow of information and disease spreading in social insects colonies: effects of environmental events and spatial heterogeneity. J Theoret Biol 492:1–10. https://doi.org/10.1016/j.jtbi.2020.110191

Hammel B, Vollet-Neto A, Menezes C et al (2016) Soldiers in a stingless bee: work rate and task repertoire suggest they are an elite force. Am Nat 187(1):120–129. https://doi.org/10.1086/684192

Holbrook CT, Clark RM, Jeanson R et al (2009) Emergence and consequences of division of labor in associations of normally solitary sweat bees. Ethology. https://doi.org/10.1111/j.1439-0310.2009.01617.x

Iverson KE (1962) A Programming Language. Wiley, New York

Jeanne RL (1986) The evolution of the organization of work in social insects. Monit Zool Ital

Jones SM, van Zweden JS, Grüter C et al (2012) The role of wax and resin in the nestmate recognition system of a stingless bee. Tetragonisca angustula. Behav Ecol Sociobiol 66(1):1–12. https://doi.org/10.1007/s00265-011-1246-7

Jongepier E, Foitzik S (2016) Fitness costs of worker specialization for ant societies. Proc R Soc B Biol Sci 283(1822). https://doi.org/10.1098/rspb.2015.2572

Kang Y, Theraulaz G (2016) Dynamical models of task organization in social insect colonies. Bull Math Biol 78(5):879–915. https://doi.org/10.1007/s11538-016-0165-1

Kang Y, Clark R, Makiyama M et al (2011) Mathematical modeling on obligate mutualism: interactions between leaf-cutter ants and their fungus garden. J Theor Biol 289:116–127

Kang Y, Blanco K, Davis T et al (2016) Disease dynamics of honeybees with varroa destructor as parasite and virus vector. Math Biosci 275:71–92

Kao AB, Couzin ID (2019) Modular structure within groups causes information loss but can improve decision accuracy. Philos Trans R Soc B Biol Sci 374(1774). https://doi.org/10.1098/rstb.2018.0378

Koedam D, Brone M, Van Tienen P (1997) The regulation of worker-oviposition in the stingless bee Trigona (Tetragonisca) angustula Illiger (Apidae, Meliponinae). Insectes Soc 44(3):229–244

Kuznetsov YA, Rinaldi S, Gragnani A (2003) One-parameter bifurcations in planar Filippov systems. Int J Bifurc Chaos 13(08):2157–2188

Meza MEM, Bhaya A, Kaszkurewicz E et al (2005) Threshold policies control for predator-prey systems using a control Liapunov function approach. Theor Popul Biol 67(4):273–284

Moses ME, Cannon JL, Gordon DM et al (2019) Distributed adaptive search in T cells: lessons from ants. Front Immunol 10.3389/fimmu.2019.01357, www.frontiersin.org/article/10.3389/fimmu.2019.01357/full

Mosqueiro T, Cook CN, Huerta R, et al (2017) Task allocation and site fidelity jointly influence foraging regulation in honeybee colonies. R Soc Open Sci 4(8):170,344

Naug D, Gadagkar R (1999) Flexible division of labor mediated by social interactions in an insect colony: a simulation model. J Theor Biol 197(1):123–133. https://doi.org/10.1006/jtbi.1998.0862

Pratt SC (2005) Quorum sensing by encounter rates in the ant Temnothorax albipennis. Behav Ecol 16(2):488–496. https://doi.org/10.1093/beheco/ari020

Ratti V, Kevan PG, Eberl HJ (2012) A mathematical model for population dynamics in honeybee colonies infested with varroa destructor and the acute bee paralysis virus. Can Appl Math Q 21(1):63–93

Ratti V, Kevan PG, Eberl HJ (2017) A mathematical model of forager loss in honeybee colonies infested with varroa destructor and the acute bee paralysis virus. Bull Math Biol 79:1218–1253

Seeley TD (1982) Adaptive significance of the age polyethism schedule in honeybee colonies. Behav Ecol Sociobiol. https://doi.org/10.1007/BF00299306

Segers FH, Menezes C, Vollet-Neto A et al (2015) Soldier production in a stingless bee depends on rearing location and nurse behaviour. Behav Ecol Sociobiol 69(4):613–623. https://doi.org/10.1007/s00265-015-1872-6

Segers FH, Zuben LV, Grüter C (2016) Local differences in parasitism and competition shape defensive investment in a polymorphic eusocial bee. Ecology 97(2):417–426. https://doi.org/10.1890/15-0793.1

Stephens PA, Sutherland WJ (1999) Consequences of the Allee effect for behaviour, ecology and conservation. https://doi.org/10.1016/S0169-5347(99)01684-5

Strickland L, Baudier KM, Bowers K, et al (2019) Bio-inspired role allocation of heterogeneous teams in a site defense task. In: Distributed autonomous robotic systems. Springer, Cham, pp 139–151. https://doi.org/10.1007/978-3-030-05816-6_10

Tang S, Liang J, Xiao Y et al (2012) Sliding bifurcations of Filippov two stage pest control models with economic thresholds. SIAM J Appl Math 72(4):1061–1080

Tang S, Xiao Y, Wang N et al (2012) Piecewise HIV virus dynamic model with CD4+ T cell count-guided therapy: I. J Theor Biol 308:123–134

Theraulaz G, Bonabeau E, Deneubourg JL (1998) Response threshold reinforcement and division of labour in insect societies. Proc R Soc B Biol Sci 265(1393):327–332. https://doi.org/10.1098/rspb.1998.0299

Van Veen J, Sommeijer MJ (2000) Colony reproduction in Tetragonisca angustula (Apidae, Meliponini). Insectes Soc 47(1):70–75

Wang A, Xiao Y, Smith R (2019) Using non-smooth models to determine thresholds for microbial pest management. J Math Biol 78:1389–1424

Wang L, Qiu Z, Kang Y (2022) A collective colony migration model with hill functions in recruitment. Int J Bifurc Chaos 32(14):2250,213

Xiao Y, Zhao T, Tang S (2013) Dynamics of an infectious diseases with media/psychology induced non-smooth incidence. Math Biosci Eng 10(2):445

Young KD, Özgüner Ü (1999) Variable structure systems, sliding mode and nonlinear control, vol 247. Springer, Berlin

van Zweden JS, Grüter C, Jones SM et al (2011) Hovering guards of the stingless bee Tetragonisca angustula increase colony defensive perimeter as shown by intra- and inter-specific comparisons. Behav Ecol Sociobiol 65(6):1277–1282. https://doi.org/10.1007/s00265-011-1141-2